DOCUMENT RESUME

ED 286 382 FL 016 932

AUTHOR Joseph, Brian D., Ed.
TITLE Studies on Language Change. Working Papers in Linguistics No. 34.
INSTITUTION Ohio State Univ., Columbus. Dept. of Linguistics.
PUB DATE Dec 86
NOTE 171p.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Collected Works - General (020)

EDRS PRICE MF01/PC07 Plus Postage.
DESCRIPTORS Arabic; Diachronic Linguistics; Dialects; *Diglossia; English; Estonian; *Etymology; Finnish; Foreign Countries; Language Variation; Linguistic Borrowing; *Linguistic Theory; *Morphemes; *Morphology (Languages); Old English; Sanskrit; Sociolinguistics; Syntax; *Uncommonly Taught Languages; Word Frequency
IDENTIFIERS Saame

ABSTRACT

A collection of papers relevant to historical linguistics and description and explanation of language change includes: "Decliticization and Deaffixation in Saame: Abessive 'taga'" (Joel A. Nevis); "Decliticization in Old Estonian" (Joel A. Nevis); "On Automatic and Simultaneous Syntactic Changes" (Brian D. Joseph); "Loss of Nominal Case Endings in the Modern Arabic Sedentary Dialects" (Ann M. Miller); "One Rule or Many? Sanskrit Reduplication as Fragmented Affixation" (Richard D. Janda, Brian D. Joseph); "Fragmentation of Strong Verb Ablaut in Old English" (Keith Johnson); "The Etymology of 'bum': Mere Child's Play" (Mary E. Clark, Brian D. Joseph); "Small Group Lexical Innovation: Some Examples" (Christopher Kupec); "Word Frequency and Dialect Borrowing" (Debra A. Stollenwerk); "Introspection into a Stable Case of Variation in Finnish" (Riitta Valimaa-Blum); and "Diglossia in Ancient India" (Gina M. Lee). (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED286382

Working Papers in Linguistics

No. 34

STUDIES ON LANGUAGE CHANGE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. Dowty

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Edited by
Brian D. Joseph

December 1986

OSU
FL016932

The Ohio State University
DEPARTMENT OF LINGUISTICS

BEST COPY AVAILABLE

The Ohio State University

Working Papers in Linguistics # 34

Studies on Language Change

Edited by

Brian D. Joseph

The Ohio State University

Department of Linguistics

204 Cunz Hall
1841 Millikin Road
Columbus, Ohio 43210-1229
U. S. A.

December 1986

Information about Ohio State University WORKING PAPERS IN LINGUISTICS (OSU WPL)

Working Papers in Linguistics is an occasional publication of the Department of Linguistics of Ohio State University and usually contains articles written by students and faculty in the department. There are one to three issues per year. Information about available issues appears below. Nos. 1, 5, & 10 are out of print and no longer available.

There are two ways to subscribe to WPL. The first is on a regular basis: the subscriber is automatically sent and billed for each issue as it appears. The second is on an issue-by-issue basis: the subscriber is notified in advance of the contents of each issue, and returns an order blank if that particular issue is desired. To order Working Papers, please send a check made payable to "Ohio State University" to: OSU WPL, Department of Linguistics, Ohio State University, 1841 Millikin Rd., Columbus, OH 43210-1229.

OSU WPL #18, $2.00. 183 pp. (June 1975): Papers by Michael Geis, Sheila Geoghegan, Jeanette Gundel, G. K. Pullum, and Arnold Zwicky.

OSU WPL #19, $2.00. 214 pp. (September 1975), Edited by Robert K. Herbert: **Patterns of Language, Culture, and Society: Sub-Saharan Africa** contains eighteen papers presented at the Symposium on African Language, Culture, and Society, held at Ohio State University on April 11, 1975.

OSU WPL #20, $2.00. 298 pp. (September 1975), Edited by Robert K. Herbert: **Proceedings of the Sixth Conference on African Linguistics** contains twenty-seven papers presented at the Sixth Conference on African Linguistics, held at the Ohio State University on April 12-13, 1975.

OSU WPL #21, $2.50. 252 pp. (May 1976), Edited by Arnold Zwicky: **Papers on Nonphonology.** Papers by Steven Boer and William Lycan, Marian Johnson, Robert Kantor, Patricia Lee, and Jay Pollack.

OSU WPL #22, $3.50. 151 pp. (February 1977), Edited by Olga Garnica: **Papers in Psycholinguistics and Sociolinguistics.** Papers by Sara Garnes, Olga Garnica, Mary Louise Edwards, Roy Major, and John Perkins.

OSU WPL #23, $3.50. 162 pp. (December 1978): Donegan, Patricia J., On the Natural Phonology of Vowels. (OSU Ph.D. dissertation).

OSU WPL #24, $3.50. 173 pp. (March 1980), Edited by Arnold M. Zwicky. **Clitics and Ellipsis.** Papers by Robert Jeffers, Nancy Levin (OSU Ph.D. Dissertation), and Arnold Zwicky.

OSU WPL #25, $3.50. 173 pp. (January 1981), Edited by Arnold M. Zwicky: **Papers in Phonology.** Papers by Donald Churma, Roderick Goman (OSU Ph.D. Dissertation), and Lawrence Schourup.

ISSN 0473-9604

OSU WPL #26, $4.50. 137 pp. (May 1982), Edited by Brian D. Joseph: **Grammatical Relations and Relational Grammar**. Papers by David Dowty, Catherine Jolley, Brian Joseph, John Nerbonne, and Amy Zaharlick.

OSU WPL #27, $4.50. 164 pp. (May 1983), Edited by Gregory T. Stump: **Papers in Historical Linguistics**. Papers by Donald Churma, G. M. Green, Leena Hazelkorn, Gregory Stump, and Rex Wallace.

OSU WPL #28, $5.00. 119 pp. (May 1983). Lawrence Clifford Schourup, Common Discourse Particles in English Conversation. (OSU Ph.D Dissertation)

OSU WPL #29, $5.00. 207 pp. (May 1984), Edited by Arnold Zwicky & Rex Wallace: **Papers on Morphology**. Papers by Belinda Brodie, Donald Churma, Erhard Hinrichs, Brian Joseph, Joel Nevis, Anne Stewart, Rex Wallace, and Arnold Zwicky.

OSU WPL #30, $5.00. 203 pages. (July 1984). John A. Nerbonne, German Temporal Semantics: Three-Dimensional Tense Logic and a GPSG Fragment. (OSU Ph.D. Dissertation).

OSU WPL #31, $6.00. 194 pages. (July 1985), Edited by Michael Geis. **Studies in Generalized Phrase Structure Grammar**. Papers by Belinda Brodie, Annette Bissantz, Erhard Hinrichs, Michael Geis and Arnold Zwicky.

OSU WPL #32, $6.00. 162 pages. (July 1986), **Interfaces**. 14 articles by Arnold Zwicky concerning the interfaces between various components of grammar.

OSU WPL #33, $6.00. 159 pages. (August 1986), Joel A. Nevis, Finnish Particle Clitics and General Clitic Theory. (OSU Ph.D. Dissertation).

OSU WPL #34, $6.00. 164 pages. (December 1986), Edited by Brian Joseph. **Studies on Language Change**. Papers by Riitta Blum, Mary Clark, Richard Janda, Keith Johnson, Christopher Kupec, Brian Joseph, Gina Lee, Ann Miller, Joel Nevis, and Debra Stollenwerk.

The following issues are available through either: The National Technical Information Center, The U.S. Department of Commerce, 5285 Port Royal Rd, Springfield, VA 22151 (PB) or ERIC Document Reproduction Serv. (ED) Ctr. for Applied Linguistics, 1611 N. Kent St., Arlington, VA 22209.

#2. November 1968, 128 pp. (OSU-CISRC-TR-68-3). PB-182 596.
#3. June 1969, 181 pp. (OSU-CISRC-TR-69-4). PB-185 855.
#4. May 1970, 164 pp. (OSU-CISRC-TP-70-26). PB-192 163.
#6. September 1970, 132 pp. (OSU-CISRC-TR-70-12). PB-194 829.
#7. February 1971, 243 pp. (OSU-CISRC-TR-71-7). PB-198 278.
#8. June 1971, 197 pp. (OSU-CISRC-TR-71-7). PB-202 724.
#9. July 1971, 232 pp. (OSU-CISRC-TP-71-8). PB-204 002.
#11. August 1971, 167 pp. ED 062 850.
#12. June 1972, 88 pp. (OSU-CISRC-TP-72-6). PB-210 781.
#13. December 1972, 225 pp. ED 077 268.
Parts of #14. April 1973, 126 pp. ED
#15. April 1973, 221 pp. ED 082 566.
Parts of #16, December 1973, 119 pp. ED

INTRODUCTION

This volume brings together 11 papers written over the past few years by faculty and students and others connected in some way and at some time with OSU. Some of the papers, in particular those by Miller, Johnson, Kupec, Stollenwerk, Välimaa-Blum, and Lee, were originally written for graduate-level lecture courses and seminars, while the others represent independently motivated work.

The common thread holding these papers together is that they all touch on issues relevant to historical linguistics and to the descrption and explanation of language change. They effectively reflect current work being done in historical linguistics in general and moreover are representative of those aspects of historical linguistics that are considered especially important in the OSU Linguistics Department. Thus, there is a definite sociolinguistic thrust to this collection, with several papers--those by Clark & Joseph, Kupec, Stollenwerk, Välimaa-Blum, and Lee--focusing on different aspects of dialect borrowing in language change (though each with its own special perspective, e.g. Clark & Joseph's on using a particular dialect borrowing explanation in etymologizing, Lee's on the relevant evidence for the interpretation of the extent and direction of change from earlier stages of Indic, etc.) and on the relevance of social factors in the spread of linguistic innovations.

At the same time, though, the more formal side of language change, along with other--especially internal--types of motivation for change is attended to in other papers, most notably those by Nevis, Miller, and Joseph, and in the Janda & Joseph collaborative effort, a general theory of morphological change is advanced which is tested and elaborated upon in Johnson's paper.

It should be noted as well that change in virtually all levels of language is covered, including phonetic change (Stollenwerk), phonological change (Janda & Joseph, Johnson), morphological change (Nevis, and, again, Janda & Joseph, Johnson), morphosyntacic change (Miller), syntactic change (Joseph, and, again, Nevis), lexical change (Kupec, Välimaa-Blum, Clark & Joseph), and to a limited extent, semantic change as well (in Kupec's paper).

Finally, the variety of languages covered is noteworthy: Sanskrit, Middle Indic, Greek, Arabic, Old English, Modern English, Finnish, Saame, and Estonian each form the major focal points of one or more of the papers, reflecting our belief that to understand the general phenomenon, language change, one must investigate a wide range of natural languages.

I would like to thank Sung-Ae Kim, Marlene Payha, and Hideo Tomita for their assistance in the production of this volume.

BDJ 10/20/86

OSU WPL # 34

Table of Contents

Studies on Language Change

Edited by Brian D. Joseph

Page

Joel A. Nevis, "Decliticization and Deaffixation in Same: Abessive taga" 1

Joel A. Nevis, "Decliticization in Old Estonian" 10

Brian D. Joseph, "On Automatic and Simultaneous Syntactic Changes" 28

Ann M. Miller, "Loss of Nominal Case Endings in the Modern Arabic Sedentary Dialects" 56

Richard D. Janda and Brian D. Joseph, "One Rule or Many? Sanskrit Reduplication as Fragmented Affixation" . . 84

Keith Johnson, "Fragmentation of Strong Verb Ablaut in Old English" 108

Mary E. Clark and Brian D. Joseph, "The Etymology of bum: Mere Child's Play" 123

Christopher Kupec, "Small Group Lexical Innovation: Some Examples" 127

Debra A. Stollenwerk, "Word Frequency and Dialect Borrowing" 133

Riitta Välimaa-Blum, "Introspection into a Stable Case of Variation in Finnish" 142

Gina M. Lee, "Diglossia in Ancient India" 151

OSU WPL 34 (1986) 1-9

Decliticization and Deaffixation in Saame: Abessive taga*

Joel A. Nevis

1. Introduction

Agglutination is an extremely common diachronic process in the languages of the world. As one of the oldest and best-known theories of the genesis of affixes, it has been used widely as a method of reconstructing constituent order, as in Givon's (1971) now well-known slogan "Today's morphology is yesterday's syntax." According to the agglutination hypothesis, affixes are historically former words that have lost their independence and have cliticized onto a neighbor, only later to fuse into the host as an affix, as in (1).

(1) WORD > CLITIC > AFFIX

There have been recent attempts to constrain and even refute this methodology, e.g. by Anderson (1980) and Comrie (1980) among others. In general, though, linguists have accepted the agglutinative cycle of words, even if only as a general tendency.

There are only a few instances of the opposite direction of change in the literature on agglutination (e.g. Jeffers and Zwicky 1980, Janda 1981), [illegible] an affix has become a clitic or a clitic has become a word:

(2) AFFIX > CLITIC > WORD

I adopt some relevant terminology from Jeffers and Zwicky (1980). The reanalysis of a word as a clitic is called cliticization and the reanalysis of a clitic as a word decliticization. The reanalysis of a clitic as an affix is affixation and that of an affix as a clitic is deaffixation:

(3) Cliticization: WORD > CLITIC
Decliticization: CLITIC > WORD
Affixation: CLITIC > AFFIX
Deaffixation: AFFIX > CLITIC

In Saame (Lappish) deaffixation and decliticization are possible developments. The abessive morpheme is traditionally viewed as a case ending, but I will argue that it is in fact a clitic in most varieties of N. Saame and Kildin Saame, and a full word in the Enontekiö variety of N. Saame (section 2). Afterwards I will demonstrate that the abessive originates historically as an affix (section 3 and 4).

2. Evidence for Synchronic Loose Status

The motivation for the clitic postposition status of abessive taga [haga] comes from its syntactic properties. It exhibits the

syntax and morphology of regular postpositions in that it governs the genitive on the preceding noun phrase, it permits "conjunction reduction", it attaches outside possessive enclitics, and it disallows adjective-noun concord . A typical case suffix, in contrast, attaches to a stem rather than a fully formed genitive noun phrase, does not permit "conjunction reduction", attaches inside possessive enclitics, and allows adjective concord wherever appropriate.

The paradigm in (4) shows the morphotactics of a noun. Notice that case morphemes precede possessive morphemes. This is entirely expected since the case endings are generally[1] true suffixes and the possessives are enclitics (Collinder 1949: 7, 1957: 193-94) - clitics always attach externally to affixes when the two cooccur (cf. Zwicky 1977).

(4) POSSESSIVE PARADIGM for ak'ko 'grandmother'
plus possessive -m 'my' (from Itkonen 1960:46-49)

	SINGULAR
NOMINATIVE	ak'ku m
GENITIVE	ak'ku-m
ACCUSATIVE	ak'ku-m
ILLATIVE	ak'ku sâ m
LOCATIVE	akko stâ-m
ESSIVE	ak'ko nâ m
COMITATIVE	akko-inâ m

	PLURAL
NOMINATIVE	akko-idâ-m
GENITIVE	akko idâ-m
ACCUSATIVE	akko-idâ-m
ILLATIVE	akko-idâ sa m
LOCATIVE	akko-i nâ m
ESSIVE	akko nâ m
COMITATIVE	akko-idâ-m-guim

(5) MORPHOTACTICS of the Locative Plural

akko-(idâ) -sa -m 'to my grandmother'
STEM (NUMBER) CASE POSS

In those dialects and languages that permit possessive plus abessive at all, the abessive noun phrase has the opposite ordering, with possessives preceding the abessive morpheme:

(6) MORPHOTACTICS of the Abessive (Ter dialect of Kildin Saame as reported by Szabó 1984: 168)

alaga-n-ta 'without my son'
son-1SG-ABES

Either (6) displays an endoclitic possessive -n or as I argue here the abessive is not a true suffix.

Although adjective-noun concord is very limited in Saame, it does appear in certain demonstrative and interrogative pronouns, in cardinal numbers and in the adjectives buorre 'good' and bâhâ 'bad'. True case suffixes exhibit agreement, e.g. in (7), but postpositions do not. The abessive patterns with the postpositions in this regard, as in (8) below.

(7) AGREEMENT -- Locative Plural (Korhonen 1981:345)

bûrii-n mâni-i-n 'to the good children'

(8) NO AGREEMENT in Abessive (Sammallahti 1983: 174)

čiežaid gâpmagii-haga 'without seven shoes'

*čiežai-haga gâpmagii-haga

No case suffix permits "conjunction reduction" (to use transformational terminology), yet the abessive allows it, even prefers it. Compare the conjoined noun phrases in (9) and (10), where (a) represents the full versions and (b) the reduced versions. Conjunction reduction is also preferred for postpositional phrases, as in (11), where relevant postpositional phrases are bracketed for ease of exposition. Again the syntactic behavior of the abessive parallels that of the postpositions.

(9) SUFFIX -- Comitative Singular (Sammallahti 1983:56)

a. Áhčiin ja Issáhiin
father-COM and Issat-COM
'with father and with Issat'

b. *Áhči ja Issáhiin
father(GEN) and Issat-COM
'with father and Issat'

(10) ABESSIVE (Bartens 1984)

a. airoj taga ja borjas taga
oar-PL ABES and sail-ABES
'without oars and without a sail

b. airoj ja borjas-taga
oar-PL and sail-ABES
'without oars and a sail'

(11) POSTPOSITIONS (Bartens 1978: 17,77)

a. ja ruohta [alemus varidi nala] ja [kaisait nala],
and runs highest mountains up and summits up

kos lae kalosaebbo.
where is cooler

'and runs up the highest mountains and up the summits where it is cooler.'

b. ja ruhtet [cuoikait ja bahka sivas] ala vare
and run mosquitos and heat reason high mountain

luokait.
along

'and (they) run because of the mosquitos and heat
along the high mountain'

Finally, in all of the above examples the abessive governs the genitive on the preceding noun phrase. This is typical of regular postpositions in the language. By comparison, true case suffixes either have concord between appropriate modifier and head noun, or else require a (default) attributive form for the modifier.

In summary, the abessive behaves synchronically like a postposition and thus should not be considered a case suffix. The only property distinguishing _-taga_ and other postpositions is its attachment to the preceding noun. The nature of this attachment is not clear to me at this point -- there is no word-internal phonological evidence to prove that the abessive attaches phonologically. Therefore it does not deserve to be called a clitic. Instead, it appears that _-taga_ is simply a stressless postposition, which cannot accept stress under any circumstances (Sammallahti 1983). I conclude that abessive _taga_ is synchronically a semi-clitic postposition.[2]

In the Enontekiö dialect of Northern Saame, the abessive has complete phonological independence and is not even a semi-clitic (much less clitic) postposition:

(12) Western Enontekiö: mun báhcen haga 'I remain without'
Eastern Enontekiö: mun báhcen taga 'I remain without'
I go without

In these parallel examples, _haga_ and _taga_ appear as adverbs and do not require a host for cliticization or prosodic learning.

3. Comparative Evidence for Former Affix Status

To return to the historical side of the topic, it could be argued that Saame retains what was in Proto-Finno-Ugric a full word, which in all other Finno-Ugric languages underwent cliticization with subsequent affixation. This is certainly a plausible parallel development, considering that most Finno-Ugric languages are agglutinative by nature. But comparative evidence uncovers certain idiosyncrasies in nearly all the sister languages, demonstrating that the abessive morpheme should be reconstructed as an affix in the parent language.

Most of the Finno-Permic languages have cognate abessive case endings (13); the Ugric branch exhibits a somewhat different cognate set. The proto-form from which these Finno-Permic abessive are derived is caritive *_pta_ plus lative *_k_.

(13) ABESSIVE CASE ENDINGS (Bartens 1984)

Southern Saame	-pta, -t'ta
Northern Saame	-t'ta, -taga
Finnish	-tta', -ttä'
Mordva	ø
Mari	-tə̑, -ðə̑
Udmurt	-tek
Komi	-teg

Northern Saame taga displays a pleonastic lative *-k/-n (Korhonen 1981, Bartens 1984):

(14) taga < *pta+k+(e)k/n (Caritive *pta + Lative *-k + Pleonastic Lative *-k/-n)

This caritive element appears also in an adjectival suffix -- in the Finno-Permic languages it is formed with a -ma suffix (15a); in the Ugric languages it has an *-l suffix (15b).[3]

(15) ADJECTIVAL CARITIVES (Bartens 1984)

a.

N. Saame	-tabme
Finnish	-ton, ttoma-
Mordva	-vtomo, -vteme; -ftama, -ftəma, -ftə́mə
Mari	-təm, -ðəm
Udmurt	-tem
Komi	-teg

b.

Hanti	(-tam, -tem, borrowed from Komi)
Mansi	-tal, -täl
Hungarian	-talan, -tlan, -telen, -tlen

It is thus unlikely that *-pta was an independent word which cliticized in all of the sister languages except for a few varieties of Saame. And additional evidence shows that cliticization would have had to take place at an early stage in the development of the Finno-Permic languages. Most of these languages exhibit relic verbal abessives in which the nominal abessive (*-pta+k) is attached to the verbal base, as in (16) below. Since case affixes do not generally attach to verb stems, or else require an intervening nominalizing suffix, the relic forms are unpredictable and, again, not a likely parallel development. Therefore the bound nature of the *-pta morpheme is due to genetic inheritance from the Finno-Ugric parent language.

(16) VERBAL ABESSIVE (Bartens 1984)

N. Saame	(taebme)
Finnish	-tta', -ttä'
Mordva	ø
Mari	-de
Udmurt	-tek
Komi	-teg
Hanti	-ləɣ, -ɲiək, -ɲəɣ

The comparative method dictates that we reconstruct a (derivational) suffix *-pta which in conjunct with lative *-k formed an abessive case ending and which in combination with derivational suffix *-ma (or *-l for Ugric) formed a caritive adjectival suffix. Thus I posit the following development:

(17) LATE PROTO FINNO-UGRIC

The methodology requires that the source for Northern Saame be a suffix and that deaffixation and decliticization be innovations in Northern and Kildin Saame.

4. Language-Internal Evidence for Former Affixal Status

Language-internal evidence also indicates that the independence of taga is an innovation. In most dialects of Northern Saame there is another abessive allomorph, namely -t'ta (Collinder 1957: 190, Sammallahti 1983:167-68), which does not have the word-like characteristics of taga. The -t'ta allomorph occurs with trisyllabic stems, as in (18).

(18) gabmâsâ t'ta 'without a (reindeer) skin'
(Collinder 1957: 190)

The taga allomorph occurs with stems having an even number of syllables:

(19) dolâ tâgâ 'without fire' (Collinder 1957: 190)

Although the two are in complementary distribution, the taga allomorph is now spreading at the expense of the -t'ta allomorph. This is allowed because of stem allomorphy in the paradigm, whereby stem allomorphs can alternate according to number of syllables, as in (20).

(20) gabmâsâ t'ta — gabmâs tâgâ 'without a (reindeer skin)'
(Collinder 1957:19)

5. Summary

The scheme that I offer here requires that a former affix sequence *-pta-k(-k/-n), which formed that abessive, have acquired a certain amount of phonological independence in several varieties of Saame and later (in Enontekiö) complete independence; these developments are illustrated in (21):

(21) -pta k ek/n > -ptakek/-ptaken > -taga > taga/hâga
AFFIXES AFFIX CLITIC WORD

Northern Saame has a semi-clitic abessive -taga which used to be a true affix. In the Enontekiö subdialect it has come to stand as an independent word. I have explained that these two taga morphemes have come about through, first, deaffixation and, then, decliticization. The Saame data discussed here constitute a good example of the opposite of agglutination: bound forms can acquire phonological independence to become independent words. I conclude that "today's syntax can be yesterday's morphology."[4]

Notes

* This paper was composed at the Ohio State University, though I prepared the final version while on the faculty of the University of Michigan. The paper was presented to the 1985 LSA Winter Meeting in Seattle. Transportation to that meeting was provided in part by funding from the OSU Linguistics Department. The ideas expressed here in are the result of research carried out in 1983-84 while on fellowship at the University of Helsinki (Supported by the American-Scandinavian Foundation and by the United States Educational Foundation in Finland). I benefited greatly from the input provided by University of Helsinki Professors Raija Bartens and Mikko Korhonen, and I hereby express my gratitude to them.

1. The morphotactics of comitative plural guim reveal that it, too, is a clitic postposition rather than a true suffix. See Nevis (1986ms).

2. A semi-clitic is a prosodic leaner, and does not interact phonologically with the host. A veritable clitic has clear phonological interaction with the host. See Nevis (1985a,b) for a discussion of the relevant terminology with respect to Finnish and Estonian.

3. In the discussion at the LSA meeting, Robert Austerlitz suggested that the -ma suffix has the same etymology as the deverbal suffix -ma of Finnish. This would mean that the *pta + k combination (i.e. derivational plus inflectional suffix) merged into a case ending, while *pta + ma (two derivational suffixes) fused into a single adjectival suffix. If *pta was a derivational suffix, then the relic verbal abessive forms are no longer a mystery. Although inflectional affixes are generally selective in stem selection, i.e. limited to a single word class, derivational affixes can sometimes be promiscuous in stem selection, relying on semantic rather than syntactic information

4. This is Janda's paraphrase of Givon's paraphrase cited above in the first paragraph. However, neither Janda's data nor my own prove that deaffixation and decliticization have any effect on constituent order.

References

Anderson, Stephen. (1980). On the development of morphology from syntax. In Historical Morphology, edited by Jacek Fisiak. (Trends in Linguistics: Studies and monographs 17.) Mouton, the Hague.

Bartens, Raija. (1984). Synteettiset ja analyyttiset rakenteet Lapin paikanilmauksissa. SUST 166.

Bartens, Raija. (1984). Uralilainen muoto- ja lauseoppi. Lectures, University of Helsinki.

Collinder, Björn. (1949). The Lappish Dialect of Jukkasjärvi. Almqvist & Wiksells, Uppsala. (= Skrifter utgivna av K. Humaniska Vetenskaps-Samfundet i Uppsala 37.3.)

Collinder, Björn. (1957). Survey of the Uralic Languages. Almqvist & Wiksells, Uppsala.

Comrie, Bernard. (1980). Morphology and Word Order Reconstruction: Problems and Prospects. In Historical Morphology, edited by Jacek Fisiak. (Trends in Linguistics: Studies and Monographs 17.) Mouton, the Hague.

Givón, Talmy. (1971). Historical Syntax and Synchronic Morphology: An Archaeologist's Field Trip. CLS 7: 394-415.

Itkonen, Erkki. (1960). Lappische Chrestomathie mit grammatikalischem Abriss und Wörterverzeichnis. Hilfsmittel für das studium der finnisch-ugrischen Sprachen VII. Suomalais-Ugrilainen Seura, Helsinki.

Janda, Richard. (1981). A Case of Liberation from Morphology into Syntax: The Fate of the English Genitive Marker -(e)s. Syntactic Change. Natural Language Studies 25: 59-114.

Jeffers, Robert, and Arnold Zwicky. (1980). The Evolution of Clitics. Papers from the Fourth International Conference on Historical Linguistics. edited by E.Traugott et al. John Benjamins. Also in OSU WPL 24: 55-64.

Korhonen, Mikko. (1981). Johdatus lapin kielen historiaan. SUST 370, Helsinki.

Nevis, Joel. (1985a). Finnish Particle Clitics and General Clitic Theory. Ph.D. dissertation, the Ohio State University.

Nevis, Joel. (1985b). Current Topics in Estonian Clisis. Paper presented to the Finno-Ugric Studies Association of Canada, Montreal.

Nevis, Joel. (1986). A Problem in Northern Lappish Morphology. Manuscript.

Szabó, László. (1984). The Functions of the Essive, Comitative, and Abessive in Kola-Lappish. Nordlyd, Tromsø University Working Papers on Language and Linguistics 9: 156-71.

Sammallahti, Pekka. (1977). Norjansaamen Itä Enontekiön murteet äänneoppi. SUST 160.

Sammallahti, Pekka. (1983). Saamen peruskurssi 1 & 2. Kokkeilumoniste, second edition. Lapin Sivistysseura ry, julkaisu 45.

OSU WPL34 (1986) 10-27

Decliticization in Old Estonian*

Joel A. Nevis

1. Introductory Remarks

Agglutination is a universal diachronic process among the languages of the world. As one of the oldest and best-known theories of the genesis of affixes, it has been used widely as a method of reconstructing word order -- as in Givon's (1971) well-known slogan "Today's morphology is yesterday's syntax." Givon's methodology has been constrained by some (e.g. Anderson 1980, Comrie 1980, and others) and has been refuted by yet others. In general, though, linguists have accepted the agglutinative cycle of words, whereby affixes are historically former words which have lost their independence and cliticized onto a neighbor, inevitably later fusing into the host as an affix.

According to Givon's principle, an affix continues the positioning within the word that its word source had within the phrase. Comrie points out by way of criticism that some clitics exhibit a special attraction to the position after the first constituent of a clause -- a positioning not shared by full words. However, I have argued that these clitics are phonologically dependent syntactic words (Nevis 1985a); in addition, examples of full words occupying this "second position" slot are not uncommon (Nevis 1985a, Wackernagel 1982). The clitics in question are generally sentential in scope (Kaisse 1985), and are called 'second position' or 'Wackernagel-type' clitics. Second position clitics have a peculiar resistance to completing the agglutination cycle, so that Comrie's remarks are not to be rejected out of hand after all.

In Baltic Finnic one finds several Wackernagel-type clitics that have been diachronically stable. Interrogative *ko, informal *s, and emphatic *pa exist in most BF languages today as clitics, and continue that status from the parent language, Late proto-BF. In Estonian these morphemes have been lost as clitics, but instead of becoming affixes, they have either decliticized into seperate words or disappeared altogether.

1.1. On the Agglutination Hypothesis

I adopt here the terminology of Jeffers and Zwicky (1980). By cliticization I mean the reanalysis of a word as a clitic. Decliticization is the reverse metanalytic reinterpretation of a clitic as a word. Affixation is intended here to be a diachronic process: the reanalysis of a clitic as an affix. Conversely, deaffixation is the change from a former affix to a clitic.

(1) a. WORD > CLITIC > AFFIX
(cliticization; affixation)

(2) a. WORD < CLITIC < AFFIX
(decliticization; deaffixation)

Examples of the agglutination cycle (i.e. (1) above) are numerous. Examples of loosening (i.e. (2) above) are rare. One such example can be found in Janda (1981), who examines the history of the English genitive marker '<u>s</u>, arguing that it is a clitic in Modern English with a source in Old English as an affix.

Janda argues that deaffixation in this instance came about in Middle English through the homophony of genitive <u>-es</u> and unstressed pronoun <u>(h)ys</u> 'his', e.g. <u>my moder ys sake</u> (see Janda 1981 for more examples and details of the analysis). Janda also suggests that, had it not been for this homophony and deaffixation, Middle English would have lost genitive <u>-es</u> along with all the other case endings (as has happened in the Northern British dialects — Janda 1981:fn.4). Janda's analysis is not uncontroversial;[1] Carstairs (1985), for example, has an alternative account of deaffixation in Middle English.

Jeffers and Zwicky (1980) likewise offer an analysis from reconstructed particles in Proto-Indo-European that putatively falsifies the "tacit assumption that that clisis is invariably one stage in an inexorable development toward the status of an affix or toward ultimate oblivion". Actually there is no evidence to prove the clitic status of the particles in their example -- Second Position cannot be used as an indication of clitichood here (Nevis 1985a). Even Wackernagel (1892) acknowledged the availability of this slot for unstressed words in general (what he called quasi-enclitics).

Nevis (1985b) offers an instance of both deaffixaton and decliticization of an inflectional morpheme in Northern Lappish.

There are other examples of deaffixation and decliticization in the literature, but these usually involve changes in the lexicon (i.e. they are derivational morphemes). Several these examples are discussed by Vesikansa (1977) for Finnish. A common example in many parts of Europe is the decliticization of <u>-ism</u> (see Ariste 1968-69). In English, Finnish, and Estonian, one can speak of all kinds of "isms", with <u>ism</u> itself having the meaning 'doctrine, theory'.

As a reaction to the Agglutination theory of the origin of affixes, Tauli (1953) tells us, Alfred Ludwig[2] postulated the Adaption theory, and later Jespersen (1922) the Secretion theory. Both theories entail metanalytical reanalysis. Underlying Ludwig's Adaptation theory is a reanalysis of derivational or emphatic elements as inflectional. Jespersen's Secretion theory involves a reanalysis of "one portion of an originally indivisible word as coming to acquire a grammatical significance" (1922: 77). The possibility of metanalytical reinterpretation in morpho-syntactic change also permits the change from bound morpheme (i.e. affix or clitic) to full word.

Old Estonian offers a further example of decliticization. In Old Estonian two Second Position Clitics, namely interrogative <u>es</u> and emphatic <u>ep</u>, are free words showing no phonological interaction with a preceding word. Following Ariste (1973) and Alvre (1976, 1981), I argue that these two words have their source in Proto-Baltic Finnic Second Position clitics *<u>s</u> and *<u>pa</u>.

1.2. <u>On Clitics</u>

A clitic is a morpheme (possibly morphologically complex) having a mixed word/affix status. This is to say that it has some properties of words and some properties of affixes. Both diachronically and synchronically the clitic appears to be intermediate between the word and the affix.

Continuing along the lines of previous work (Nevis 1985a), I adopt the position that clitics are not primitive units of languages. Rather they fall into two classes of derived phenomena: (phonologically) bound words and phrasal affixes. The former is a special kind of word, the latter a type of affix. The bound word, in particular, has the syntax of words, but the phonological properties of affixes.

The Finnish clitics -pa/-pä, -ko/kö, and -han/-hän are bound words, as argued by Nevis (1985a). Although positioned with respect to the sentence as a whole, these morphemes cannot stand alone as independent words. Their phonological subordination to a preceding word is demonstrated by the application of vowel harmony. Vowel harmony operates over the phonological word, never beyond. Thus it can be used as a test for clitichood in Finnish:

tuuli-han 'the wind, you know' -- *tuuli-hän
tyyli-hän 'style, you know' -- *tyyli-han

The Baltic Finnic clitic,as represented by Finnish, are Wackernagel-type, or Second Position (2p), clitics. These clitics are bound words that, in an otherwise free constituent order language, occur strictly positioned after the first constituent of the clause and enclitic on it. See section 2.1. for examples.

Second Position clitics (or actually Second Position bound words) have a particular resistence to affixation -- both to phrasal affixation and to regular affixation. I demonstrate this point with data from Old Estonian. The Proto-Baltic Finnic Second Position clitics inherited by Estonian show some instability, namely decliticization. On the basis of the Baltic Finnic data, I offer a general account of possible and probable developments for Wackernagel-type clitics.

2. Ep and Es in Estonian

There are no Second Position clitics in Modern Estonian. From the point of view of her sister languages, a gap exists in Estonian. To account for this gap, one assumes that decliticization has taken place only in Estonian. This analysis is supported by two kinds of evidence. There is first an argument based on complementarity -- where the other Baltic Finnic languages have 2p clitics, Old Estonian has full words. The second argument relies on shared relic forms in all the Baltic Finnic Languages. Some support comes from a third source -- relic forms in Estonian alone.

2.1. Wackernagel's Law

Cognate morphemes in sister languages are Second Position, or Wackernagel-type, clitics. That is, they are phonologically bound words which are attached to the first constituent of a clause (no matter what that constituent may be). Corresponding to Estonian emphatic ep are emphatic clitics in the various sister languages:

(3)	Finnish	-pa/-pä	Tule-pa kotiin
			come-EMP home
	Karelian	-bo	midä-bo
			what EMP (Ahtia 1936:9)
	Lyydi	-bo	midä-bo
			what EMP (Larjavaara 1979:109)

Similarly, the BF cognates to Estonian interrogative *es* include Finnish, Karelian, Lyydi, Vepsä *-s*, which indicates not interrogativity, but informality (i.e. *-s* is a register marker). I account for the semantic discrepancy below in section 3.4.

(4) Finnish -s: tule-s tänne
come IF here

Karelian -s: ottakkoa-s 'take (2PL IMP)'
take(2PL) IF (Ahtia 1936: 132)

These are a subset of the various BF sentential clitics that obey Wackernagel's law. Other such clitics include BF *-ko*, northern BF *-han/-hän*, Lyydi-Vepsä *se*, and a few more clitics (Penttilä 1957, Ahtia 1936).

(5) Finnish -han/-hän anna-han mennä 'let her/him/it go'
give-HAN go

Votic -ko evät ko sȫ̂ 'don't they eat?'
not-Q eat (Ahlqvist 1858:5)

Lyydi se kod'ii se ruoh'tinu tulda en
home EMP dared come not
'Home I didn't dare (to) come'
(Larjavaara 1979: 116-17)

On the basis of comparative evidence, we want to reconstruct for Proto-Baltic Finnic at least three Second Position clitics: emphatic **pa*, informal **s*, and interrogative **ko*. The other clitics are more recent innovations. There are few traces of these clitics in Modern Estonian. For this reason one looks to *ep* and *es*, which correspond roughly in meaning and positioning, as continuers of the clitics.

Beyond the evidence presented in the following section, it is not entirely clear that **pa* and **ko* were true clitics and not just quasi-enclitics (i.e. stressless words) in the parent language. It turns out that their clitichood has no bearing on the decliticization analysis I present below, since clitichood can be established for Old Estonian on the basis of internal reconstruction.

2. Estonian *ep*

Ariste (1973) contends that Estonian affirmative emphasis marker *ep* is a direct continuation of Proto-Baltic Finnic clitic **pa*. It appears in roughly the same sentential slot as *-pa* in other BF languages, and it has the same meaning. Ariste cites a number of examples from turn-of-the-century Estonian and from the modern dialects. I repeat select ones below; for a more complete listing I refer the reader to Ariste (1973).

Emphatic *ep* is generally located after the first constituent of the sentence, as in the following examples:

(6) selle kivi peal ep kolgitigi neid riideid
this rock on EMP pounded these clothes
'on this stone one pounded the clothes'
(Ariste 1973:33; originally from Saareste 1958)

(7) Mis ep saame näha
what EMP get see
'What do we get to see'

(8) Need ep vist sinu omad ongi?
these EMP probably your own is-EMP
'These are probably your own?'

(9) Siis ep ilmub Isakene, Tuleb taevast taadikene.
then EMP appears daddy comes sky-EL granddaddy
'Then Daddy appears, granddaddy comes from the sky.'

(10) See ep siis tuli, et temal veel tihli sõbradele oli anda
there EMP then came that him still often friends-AL was give
'There then it came that he still often had a gift for friends'

(11) See ep see on, mis suurem rahvahulk,
that EMP that is what larger crowd

et vanal viisil mõtleb, ära mõista ei või
that old-AD way-AD thinks understand not can

'That is that, what a larger crowd that thinks in the old way cannot understand'

Most examples involve a one-word constituent at the beginning of the clause, but examples like (6) above show that multiword constituents can occur before ep as well. The location of ep in second position is not strict, however, as demonstrated by (12).

(12) Täna näeb ta kirikus Hildat ja temale ta ep lilled
today sees he church-IN H. and her-AL he EMP flowers
viibki
brings-EMP
'Today he sees Hilda in church and to her he brings flowers'

Thus, ep appears in the "third" slot in this example. Ariste further suggests that it can occur in sentence-initial position: see ep > ep see.

Not only has ep acquired phonological independence in its development from Proto-Baltic Finnic, but it seems to have acquired a certain amount of syntatic freedom as well. Numerous relic forms remain in the modern language, so that we know that ep was a bound morpheme. These are discussed below in section 4.2.

2.1. Negative ep

Ariste points out that Wiedemann (1857) considered (e)p a clitic in his dictionary, but confused it with the negative verb ep. The old third person singular of the negative verb was ep; it has been replaced in modern Estonian by ei, which has now spread to all persons and numbers. Nonetheless one still finds in the modern language ep ole (= ei ole) 'is not' and ep olnud (= ei olnud) 'was not' alongside their proclitic forms pole and polnud.

The continuation of Proto-Baltic Finnic *pa is affirmative emphatic ep, not negative verb ep.

3. Estonian es

Modern Estonian exihibits several es words, only one of which is a direct continuation of Proto-Baltic Finnic *s. The various homophonous morphemes are discussed below.

3.1. Interrogative es

Old Estonian interrogative es has been discussed by Ojansuu (1922) and Alvre (1976, 1981). It is a Second Position word, as seen in the 1686 New Testment examples that follow (from Ojansuu 1922: 93-94).

(13) Nüüd es tee uSSute (Joh. 16: 31)
now Q you believe
'Now do you believe?'

(14) Kelt es Se Proweet Seddä ütlep (ApT. 8: 34)
whom-ABL Q the prophet that says
'About whom does the prophet say that?'

(15) Kumb es Se Suurem KäSk om SäduSSen (Matt. 22: 36)
Which Q the larger order is law-ILL
'Which is the greater command under/according to the law?'

(16) MiSt es minna Seddä peä tundma (Luk. 1: 18)
What-EL Q I that must know
'How should I know that?'

(17) Mink ka es Sis Soolata (Matt. 5: 13)
What also Q then salt
'What kind then should you be?'

(18) märätse es teije Sis peäte ollema (Joh. 6: 30)
what-kind Q you then must be
'What kind then should you be?'

(19) Märäst Tähte es Sinna teet ...? (Joh. 6: 30)
What-kind letter Q you make
'What kind of letter/mark are you making ...?

(20) märätsel Nimmel es teije Sedda ollete tennu? (ApT. 4: 7)
what-kind name Q you that are done
'In whose name have you done that?'

(21) Kes om minno Emmä, ninck kumma es omma minno Welle
who is my mother and which Q is my brother
'Who is my mother, and which is my brother?' (Matt. 12: 48)

Examples (17-19) demonstrate that location in the sentence is determined by constituents, not words, since interrogative es appears not after the first word of the clause in these examples, but after the first constituent. One example shows that, like emphatic ep, es can occur elsewhere in the sentence than in Second Position.

(22) Kellega teie es sĩn tahate kõnelda
whom COM you Q here want speak
'With whom here do you want to speak?'
(from Wiedemann's Grammar, cited by Alvre 1976: 346)

Several of early examples still show es as a clitic. The word cannot maintain any phonological independence since there is no vowel present. All s examples are written together with the preceding word (23-24), so that we have orthographic evidence that es was once a Second Position Clitic.

(23) -- Kustas meije Lanen ni paljo Leiba Same (Matt. 15: 33)
whence Q our kind get so much bread get
'-- from where does our kind get so much bread?'

(24) Sinnas ütsinda wõras ollet... (Luk. 24: 18)
you Q alone stranger are
'Are you alone the/a stranger...?

In both of these examples, the final s of the first word (Kustas and Sinnas) is the interrogative clitic.

There is a semantic discrepancy in the correlation between Estonian interrogative es and BF informal -s. I deal with this problem below in section 3.4.

3.2. Conjunction es

Interrogative es is homophonous with, and according to some, etymologically identical to conjunction es. The latter is found in a seventeenth century Southern Estonian verse, as well as in Old Finnish (that is, the southwest dialect used by Agricola, Finno, and Hemminki). In both Southern Estonian and Old Finnish, Ojansuu (1922) tells us, es had the meaning 'if, although'. Ojansuu's contention that 2P interrogative es originated in initially positioned conjunction es is a viable alternative account to Alvre's, so it must be examined in detail.

In all of Ojansuu's examples, conjunction es begins the clause, as conjunctions generally do in Finnish and Estonian.

(25) South Estonian:
Es mejie juhren olles Sedda pattu ni palju
if our place being that sin so much

jummala juhren om wehl ennamb armu.
god's place is still more favored

'If in our place was so much of that sin, God's place is even more favored.'

(26) Agricola:
Eipe heiden pidhe woittaman ..., Es quinka corkiaSti
not-and they must win although how highly

he lendeuet: esquiga he ennen cucoiStit
they fly although-how they before flourished

'And they must not win ..., No matter how high they fly;
— no matter how they flourished before.'

(27) Finno:
es cuSa Inhiminen olis
if where person would-be
'if a person would be somewhere'

(28) Hemminki:
Es cuca vihans pääh&n nacka
if who anger head-ILL flings
'If someone flings his anger into his head'

Conjunction es is derived from a former demonstrative pronoun/relative/indefinite pronoun stem *e-. The e-stem also occurs in the Finnish conjunctions että 'that", ellei 'if not', and dialectal ellä 'if', and in Estonian emb kumb 'either (one or the other)', cf. Hakulinen (1979: 74). Paasonen (1906) finds cognates of the Baltic Finnic e-system in Mordvin, Zyrian, Votyak, Ostyak, and Hungarian, and suggests that the e-stem is a variant of the jo-stem (cf. Finnish jo-s 'if' just like conjunction e-s).

The s in conjunction es is a lative ending, which is also found in the Finnish conjunctions jos 'if' and koska 'because, when' (the -ka here is a former clitic -- see Nevis 1984), and in the adverbs alas 'down', ylös 'up' and pois 'away'.

Ojansuu proposes that conjunction es is the source for interrogative es. His proposal would require that there was a change in positioning in the sentence, a semantic change 'if, although' > 'whether', and a shift from dependent clause to main clause. It would leave open the question of why there is an absence of 2P clitics in Estonian and it would leave unexplained all the relic forms in Estonian (cf. section 4 below). Ojansuu's account would have to posit not only the two syntatic changes and the semantic change, but also a phonological change (enclisis-- examples (23) and (24) above have clitic -s).

Since some dialects have both interrogative and conjunction es words, Ojansuu would also have to assume a syntactic and semantic split.

3.3. Past Tense Negative es

Interrogative es is also homophonous with another unrelated form in the language, namely negative past tense es. Some of the Estonian dialects have innovated a past tense for negative verb (stem in e-). In general in Baltic Finnic, the negative verb e- has a defective paradigm. It may be inflected for person and number, but not for mood and tense. The Estonian dialects have allowed the negative paradigm to include tense, so that e-p is present tense and e-s is past tense, just like laula-b 'sings' and laula-s 'sang'.

(29) es näe' timä miDaGi' (Savijärvi 1981: 111)
not past see s/he something
'She/he did not see anything'

3.4. Interrogative Negative es

A third homophonous es comes from the negative verb e- plus clitic -s or from clitic combination -ko-s. As in Finnish, the negative verb ei combined with clitics *ko and *s to form a single lexicalized unit: es < *ei-ko-s. (Finnish has eikös, eiks.) The difference between the affirmative and negative interrogatives can be seen in the morphology of the following verb and in the positioning of the particle es. If the main verb is marked for person and number, then the cooccurring es is affirmative; if the main verb is not marked for person and number, then the cooccurring es is negative. The reason for this is that the negative in Baltic Finnic is an auxiliary verb and takes the person and number marks while the main verb has a special complement form. Also, if affirmative es is used, then the particle is located in Second Position. But if negative es is used, the particle is always sentence-initial. This is because the negative verb counted as the sentence initial constituent to which 2P *-ko-s attached. Relevant examples are (30) and (31). Example (30) has finite main verb om 'is' (predecessor to modern Estonian on) rather than negative complement ole (as in (31)). It further exhibits 2P es instead of initial es.

(30) Old Estonian: (= (15))
Kumb es Se Suuremb KäSk om Sädussen (Matt. 22: 36)
Which Q the larger order is law-ILL
'Which is the greater command under/according to the law?'

By contrast, the dialect example in (31) has the negative complement ole 'be' rather than main verb on 'is'. And in this example es is initial rather than second.

(31) Pältsamaa dialect (SW Estonian) from Alvre (1976: 346):
eš ta aTvem ole
not Q it cheaper be
'Isn't it cheaper'

4. Relic Forms

Evidence for the declitic analysis of Estonian ep comes from the existence of relic forms in all the Baltic Finnic daughter languages. Shared relic forms indicate that the parent language had bound words rather than free words. Further evidence lies in the large number of relic forms in Estonian itself (not shared with sister languages) showing that the sources for Old Estonian ep and es were clitic before the Old Estonian period -- pre-Estonian at the latest, most likely Proto Baltic Finnic.

4.1. Relic Forms in Baltic Finnic

All of the BF daughter languages exhibit relic forms which indicate that Second Position enclisis is inherited from the parent language. These include emphasis word juba/jopa 'even', negative plus interrogative eks/eiks/eikos, and negative plus connective egas/eikas.

According to Kalima (1936), all the BF languages have words that correspond to Estonian juba and Finnish jopa, both of which have an idiosyncratic, unpredictable meaning. The etymological source for juba/jopa 'even' is an old Germanic loanword ju, (Finnish jo) 'already' (Raun 1982: 21, cf. Gothic ju) plus emphatic *pa. The result is not semantically

compositional 'even already' but a special emphasis word. Juba/jopa has a typical syntax for a phonological word containing a Second Position Clitic; it can appear in other than initial position.

(32) from Kalima (1936: 144):

Finnish	jo	jopa
Estonian	juu	juba
Livonian	ju	juba, jub, j va

All BF languages likewise show relics of combinations informal clitic *s. Standard Finnish, for example, has en-kö-s 'don't I?', et-kö-s 'don't you?', ei-kö-s 'doesn't ?', etc. Colloquial Finnish has eiks 'doesn't ?'. Estonian has eks 'doesn't ?'. All of these come from the negative verb (stem in e-) and interrogative clitic *ko with optional informal clitic *s. The appearance of this form in Estonian is particularly surprising since it does not have the two clitics in question elsewhere.

(33) a. Finnish

1SG	en-k(ö)-s	NEG-Q-IF
2SG	et-k(ö)-s	
3SG	ei-k(ö)-s	
1PL	emme-k(ö)-s	
2PL	ette-k(ö)-s	
3PL	eivät-k(ö)-s	

b. Estonian eks < ei/ep + ko + s
NEG + Q + IF

Finally, the negative verb can combine with a former clitic conjunction *ka (see Nevis 1984) and informality marker *s. Finnish has en-kä-s 'and you don't', ei-kä-s 'and he/she/it doesn't'. Estonian has egas 'and doesn't'. Again the presence of Estonian egas is unexpected here, since it displays (a relic of) clitic *s.

(34) a. Finnish

1SG	en-kä-s	NEG-CONJ-IF
2SG	et-kä-s	
3SG	ei-kä-s	
1PL	emme-kä-s	
2PL	ette-kä-s	
3PL	eivät-kä-s	

b. Estonian egas < *ei/ep + ka + s
NEG + CONJ + IF

In summary, the complementarity of Estonian es and ep with the other BF languages' Wackernagel-type clitics suggests a common origin. The relic forms juba/jopa, eks, and egas indicate that the forms were once bound in BF. More evidence for the bound nature of these morphemes comes from strictly language-internal facts.

4.2. Estonian relics

Alvre (1981) cites a number of Estonian relics forms in which bound s continues former clitic *s or clitic combination *-ko-s. Bound s has an interrogative function here, even if only redundantly in conjunction with some of the interrogative hosts -- e.g. kuna 'when' -- kuna-s 'when'.

(35) From Alvre (1981):
kuna-s 'when'
kuida-s 'how'
palgu-s, palju-ks 'how much'
ammu-s, ammu-ks 'early?'
ilma-ks 'free?'
juba-ks, jooks 'already?'
kaua-ks, kaŭva-s 'far?'
kaugele-ks 'far?'
kuni-s 'up to what? as far as what?'
mina-ks 'me?'
sina-ks 'you?'
mitu-ks 'how many?'
muidu-ks 'otherwise?'
nõnda-ks 'like this?' thus?'
on-ks, on-s 'is?'
oli-ks 'was?'
pea-ks 'has to?'
veela-ks 'still?'
vähe-ks, vähä-s 'few?'

Dialectal forms include tuli-ks 'came?', räkkisi-ks-ma 'did I speak?', võtsi-ks-me 'did we take?', antsi-ks 'took?', miga-s 'what?', kumb-s 'which (of two)?', and ken-s 'who?' (Alvre 1981).

There are also a number of -p relics in Estonian. The list in (36) is taken from Ariste (1973: 36):

(36) see'p see on 'that's that' (lit. that EMP that is)
küllap 'certainly'
siisap 'then'
siiap 'hither'
temap 'he/she'
nondap 'thus'
samap 'same'
sinap 'you (SG)'
minap 'I'
kustap 'whence'
sealap 'there'
annap 'give!'
tulep 'come!'

The -p relics show greater variety in "host selection". -p attaches primarily to pronouns and adverbs, but can also be found connected to imperative verbs.

Former clitics *pa and *s cooccurred in some instances, as revealed in relics eps 'doesn't?' < e- negative verb plus *-pa plus *-s, nõndaps 'dann so' < nonda 'thus' plus *-pa-s, teps 'hinfort, von nun an' < te 'do!' plus *-pa-s, vastaps 'erst, soeben' < vasta 'just' plus *-pa-s (Alvre 1983).

In some instances relics -(k)s and -p have acted prophylactically to retain an apocopated vowel. For example, interrogative veela-ks 'still?' retains older final a, but veel 'still' does not; it shows the effects of apocope. And kustap 'whence (EMP)' likewise retains older final a, while kust 'whence' does not. In fact, a good number of Proto Baltic Finnic case

suffixes ended in a or ä (according to vowel harmony). These final vowels were generally apocopated, unless a clitic such as -s or -p prevented apocope.

(37)

ELATIVE	-st	>	-sta/-stä
INSSIVE	-s	>	-ssa/-ssä
ABLATIVE	-lt	>	-lta/-ltä
ABESSIVE	-l	>	-lla/-llä

Non-initial ä became e in Estonian.

The result of apocope in Estonian was that the final vowel came to be reanalysed not as part of the stem, but as part of the -s or -p morpheme. Thus, when *keltä lost its final vowel while *keltä s retained the vowel, the result was a realignment of the vowel with respect to the morpheme boundary (see also Alvre 1981):

(38) keltäs > keltes > kelt es
keltä > kelt

One would expect also as as a remnant of back vowel harmonic -a (e.g. kusta-s > kust as), but I assume that leveling was responsible for the spread of es at the expense of as. A parallel reanalysis is necessary to explain the development of ep (section 5.1.).

It is impossible to account for these relic forms in the cliticization approach -- retention of a former morpheme-final vowel cannot be reconciled with language-specific agglutination of an es or ep word. Instead, these relic forms demonstrate that independent es and ep used to be bound morphemes, and thus decliticization is required to account for their development.

5. Clitic Development in Estonian and Finnish

That ep and es are independent words arising from phonologically dependent words is clear from the preceding discussion. What remains to be accounted for is the semantic shift from informal -s to interrogative es.

On the basis of the Finnish and Estonian data, I propose a general account of possible and probable developments for Wackernagel-type bound words. Old Estonian es and ep demonstrate that decliticization is one possible course of change. Old Estonian also shows loss of former interrogative clitic *-ko. Finnish -ko/-kö and -pa/-pä reveal that clitics of this sort can be fairly stable as well. Affixation is also possible, though rare -- Finnish informal -s is one such example.

5.1. Decliticization in Old Estonian

Both es and ep were once phonologically bound words. The two decliticized at roughly the same time and in the same manner. When final vowels were apocopated circa 1250-1500 AD (Raun and Saareste 1966: 59, Kask 1972: 155), clitics *-pa and *-(ko)-s acted prophylactically in preventing apocope:

(39)

PROTO-BALTIC FINNIC		OLD ESTONIAN after apocope
a. *keltä	>	kelt
*keltä-s		*keltä s
b. *päällä	>	*pääll
*päällä-pä		*päällä p

(Recall non-initial *ä > e: note also diphthongization *ää > ea and certain degeminations.)

Once the default instance is the apocopated stem and the less common instance is the vowel allomorph before a clitic, the situation is ripe for reinterpretation. The morpheme boundary is "moved", so to speak, such that the vowel is considered part of the clitic:

(40) kelte-s kelt es
peale-p > peal-ep

And once vowel harmony is lost as a productive rule (circa 1650 AD -- Raun and Saareste 1966: 65), there is no longer any evidence that -es and -ep are phonologically dependent words rather than independent (though unstressed) words, i.e. quasi-enclitics. Since the rule that is responsible for phonological subordination consequently lacks motivation, it disappears from the language altogether, and the clitics are no longer "clitic" but independent. In other words, bound words are marked with respect to independent words, and I propose that decliticization here is an instance of the more general case in which marked become unmarked forms.

Although one would expect both es, ep and as, ap from the Proto-Baltic Finnic clitics, one finds only the e-vowel descendents. The two e/a alternates would be the continuations of former vowel harmonic alternates ä/a as explained in section 4.2. I have been assuming that the e-stem allomorphs spread at the expense of the a-allomorphs (except in instances of lexicalization -- section 4.2). Leveling of this type is confirmed by cases where neither e- nor a- sources occurred. Examples (7-9, 11, 13, 17) above, have not had an intervening vowel between the host and the clitic (or else had a different vowel).

(7) Mis ep ...
what EMP

(8) Need ep ...
these EMP

(9) Siis ep
then EMP

(11) See ep ...
that EMP

(13) Nüüd es ...
now Q

(17) Mink ka es ...
what also Q

Thus, for example, Mis ep (7) does not reconstruct with a vowel (i.e. not *misä-pä). These examples demonstrate the productivity and spread of es and ep at the expense of the a variants.

When former *s and *pa failed to decliticize in Estonian, they were lexicalized to specific lexeme and morpheme combinations. Therefore the

productive forms ep and es did not spread to every item. One finds a similar lexicalization with Finnish informal -s.

5.2. Affixation in Finnish

In Nevis (1985a) I argue that Finnish -s is not a clitic but an affix. It is no longer productive insofar as it does not combine with just anything. It can be found in four situations: it attaches to clitics -ko/-kö and -pa/-pä (but not to clitics -han/-hän 'you know, I wonder, by golly' or -kin/-kaan/kään 'also, too, neither'), it attaches to interrogative pronouns kuka 'who' and mikä 'what' and relative pronoun joka 'who'(but not to other pronouns, e.g. se 'it, that' or minä 'I'), and it attaches to imperatives like otta-kaa-s 'please take' but not other verbal moods(e.g. *otta-a-s 'takes').

This restricted distribution is uncliticlike. Affixes typically impose restrictions on host selection, but clitics do not -- they are generally promiscuous in attachment (Zwicky 1977, 1984). All occurrences of Finnish informal -s can be accounted for by assuming that the s-forms are allomorphs of the non s-forms. That is, tule-pa-s (come-EMP-IF) is not trimorphemic, but bimorphemic tule-pas, where -pas is simply the "informal" allomorph of -pa

Note that the two classes of phenomena in Finnish that allow informal -s are interrogative morphemes (interrogative pronouns plus their near cousins the relative pronouns, as well as the interrogative bound word -ko/-kö) and emphatic morphemes (imparatives and emphatic clitic -pa/-pä). The connection to the former is revealing. It shows the crucial link between Finnish informal -s and Old Estonian interrogative es.

I propose that the primary source for Estonian es was precisely this lexicalized clitic *-ko-s > -ks (> -s > es). Interrogativity originates in the former Second Position clitic *ko, but through successive stages involving lexicalization of -ko-s and upgrading colloquial -k(o)s, the meaning is now continued in es.

5.3. Clitic Loss

Proto-Baltic Finnic interrogative clitic *ko is now lost in Estonian. It played a role in the development of es, as described above, and it is found in relic forms eks, veelaks, and a few other relics (see (35) above). The loss of *ko probably came about through regular sound changes in the language. Apocope (circa 13th century -- Raun and Saareste 1966: 63) would have dropped the final vowel, hence *ei-ko > *ei-k, and loss of final *n, k, h would have dropped the now-final consonant (Kask 1972: 155-156), hence *ei-k > ei. The result is the awkward situation in which all interrogatives with former *ko become homophonous with declaratives.

Interrogative *-ko and es (now archaic and dialectal) have been replaced by kas. Kas is positioned clause-initially and is of uncertain origin. Alvre (1983: 82) attributes to L. Kettunen the suggestion that kas came from the imperative verb katso! 'look!'. But bimorphemic ka-s with relic of former interrogative clitics *-ko-s is just as likely (see Alvre 1983).

5.4. Wackernagel's Law

Bound words that occur in Sentence Second Position are oftentimes stable. They rarely complete the agglutination cycle by becoming affixes. This is

because they are frequently incompatible semantically with the host. There is often no semantic or syntactic connection between the Wackernagel-type bound word and the initially positioned constituent. Affixation cannot be ruled out completely, as evidenced by Finnish informal _s (section 5.3), but it seems to be one of the least likely developments of a Second Position clitic.

If the Wackernagel-type clitic is not stable, then it is either lost from the language (as with Proto-Baltic Finnic *-ko in Estonian) or it becomes independent (as with Proto-Baltic Finnic *-s and *-pa). Steele (1976) offers yet a different possible course of development -- the Second Position clitic can turn into another kind of clitic. In several Uto Aztecan languages, the Wackernagel-type clitic inherited from the parent attachment to the initially positioned constituent preceding it to the verb which followed it. Yaqui is an example:

(41) ?inepo ne-?a-me?ak (Steele 1976: 554)
I I -it-threw
'I threw it'

The former Second Position clitic pronoun ne no longer attaches to the first constituent in the sentence but to the following constituent, namely the verb. The former Second Position enclitic is now a verbal proclitic.

Ard (1977, 1978) presents similar data from the developments in the Slavic languages, although clitichood of the morphemes in question is not established for sure (i.e. they are likely to be leaners, i.e. quasi-clitics, rather than bound words). Wackernagel-type words are found in Czech, Slovak, Slovenian, Serbo-Croatian, Transcarpathian Ukrainian, and in dialectally in Polish. Attraction to the verb has taken place in the other Slavic languages. In Russian, Belo-Russian, Ukrainian (except for transcarpathian), and dialectally in Polish the cognate morphemes have turned into suffixes on the verb; in Macedonian and Bulgarian they are located adjacent to the verb -- after an imperative or participle, before a finite verb. Thus the alternative to Wackernagel's Law has been attraction to the head of the clause, namely the verb.

6. Concluding Remarks

Old Estonian es and ep evince the rare phenomenon of decliticization or loss, but also shows the possibility of a change in the direction of attachment. In an SVO language, the sandwiching of the [illegible] clitics between the initial constituent and the verb permits the verb to exert a syntactic and semantic pull on the clitic group, so that they attach phonologically to the verb. Verbal clitics are more compatible with the host, and consequently are more likely to complete the agglutination cycle and less likely to decliticize.

Notes

*This paper was composed at the Ohio State University, but completed at the University of Michigan. A general absence of available materials forces me to leave out some potentially relevant Karelian, Vepsä, Votic and Livonian data.

Irrelevant grammatical information is left out from glosses. Abbreviations used in this article include:

ABL	ablative
AD	adessive
AL	allative
CONJ	conjunction
COM	comitative
EL	elative
EMP	emphatic
HAN	an epistemic clitic marking reintroduced information of current discourse relevance.
IF	informal
ILL	illative
IN	inessive
PL	plural
Q	interrogative

1. Janda (personal communication) now informs me that he has given up one piece of his synchronic analysis, namely the claim that the 's morpheme is synchronically a determiner to the following NP. His diachronic analysis remains as before.

2. Tauli cites Alfred Ludwig's (1873) article "Agglutination oder adaptation?", but I have not been able to locate that reference.

References

Ahlqvist, August. 1858. Wotisk Grammatik jemte språkprof och ordförteckning. Acta Societatis Scientiarum Fennicae V.

Ahtia, E. V. 1936. Karjalan Kielioppi. Karjalan Kansalaisseura, Suojärvi.

Alvre, Paul. 1976. Vana Kirjakeele küsisõnu. Keel ja Kirjandus 1976.6: 343-350.

Alvre, Paul. 1981. Veelaks tüüp Kirjakeeles ja murdeis. Keel ja Kirjandus 1981.1: 24-30.

Alvre, Paul. 1983. Zur Herkunft der Wörter kas und teps in der estonischen Sprache. Soviet Finno-Ugric Studies 1983.2: 81-89.

Anderson, Stephen. 1980. On the Development of Morphology from Syntax. Historical Morphology. Edited by Jacek Fisiak, Mouton.

Ard, Josh. 1977. Raisings and Word Order in Diachronic Syntax. Ph.D. dissertation, UCLA (1975). Distributed by IULC.

Ard, Josh. 1978. The process of reanalysis and its implications. Paper read at the Conference on Syntactic Change, University of Michigan.

Ariste, Paul. 1968-69. Kas ism võib olla sõna? Emakeele Seltsi Aastaraamat 14-15.

Ariste, Paul. 1973. Eesti rõhumaarsona ep. Suomalais-Ugrilaisen Seuran Aikakauskirja 72:33-37.

Carstairs, Andrew. 1985. Constraints on Allomorphy in Inflexion. Distributed by the Indiana University Linguistics Club.

Comrie, Bernard. 1980. Morphology and Word Order Reconstruction: Problems and Prospects. Historical Morphology, pp. 83-96. Edited by Jacek Fisiak. Mouton.

Givón, Talmy. 1971. Historical Syntax and Synchronic Morphology: An Archeologist's Field Trip. CLS 7: 394-415.

Hakulinen, Lauri. 1979. Suomen kielen rakenne ja kehitys. WSOY, Porvoo.

Janda, Richard. 1981. A Case of Liberation from Morphology into Syntax: The Fate of the English Genitive Marker -(e)s. Syntactic Change. Natural Language Studies 25: 59-114.

Jeffers, Robert, and Arnold Zwicky. 1980. The Evolution of Clitics. Papers from the Fourth International Conference on Historical Linguistics. Also in OSU WPL 24: 55-64.

Jespersen, Otto. 1922. Language, Its Nature, Development and Origin. Holt and Co., New York.

Kalima, Jalo. 1936. Itämerensuomalaisten kielten balttilaiset lainasanat. Suomalais-Ugrilainen Seura 202, Helsinki.

Kask, Arnold. 1972. Eesti keele ajalooline grammatika. Hääldikulugu. Tartu Riiklik Ülikool, Tartu.

Larjavaara, Matti. 1979. Lyydin se-partikkeli. Sivulaudatur-työ, University of Helsinki.

Ludwig, Alfred. 1873. Agglutination oder adaptation? Prag.

Nevis, Joel. 1984. On Finnish Disyllables, Cliticization, and Typology to appear in the Proceedings of the Third Uralic Phonology Conference (Eisenstadt, Austria).

Nevis, Joel. 1985a. Finnish Particle Clitics and General Clitic Theory. Ph.D. Dissertation, The Ohio State University (to appear in Ohio State University Working Papers in Linguistics).

Nevis, Joel. 1985b. Deaffixation and decliticization in Saame: Abessive taga. Paper presented to the Linguistic Society of America Winter Meeting, Seattle. (In this volume of Ohio State University Working Papers in Linguistics).

Ojansuu, Heikki. 1922. Itämerensuomalaisten kielten pronominioppia. Turun Suom. Yliopiston Julkaisuja B.I.3, Turku.

Paasonen, Heikki. 1906-08. Die finnischen pronominalstämme jo- und e-. Finnisch-Ugrische Forschungen 6:111-17.

Raun, Alo. 1982. Eesti keele etümoloogiline teatmik. Maarjamaa, Rome Toronto.

Raun, Alo, and Andrus Saareste. 1966. An Introduction to Estonian Linguistics. Ural-Altäische Bibliothek, Wiesbaden.

Savijärvi, Ilkka. 1981. Sanajärjestystyyppi pääverbi--kieltoverbi viron kielessä. Virittäjä 85.2: 109-117.

Steele, Susan. 1976. Clisis and Dichrony. Mechanisms of Syntactic Change, pp. 539-579. Edited by Charles N. Li. University of Texas, Austin.

Tauli, Valter. 1953. The Origin of Afixes. FUF 32.1-2: 170-225.

Vesikansa, Jouko. 1977. Johdokset. Nykysuomen oppaita 2. WSOY, Porvoo.

Wackernagel, Jakob. 1892. Über ein Gesetz der indo-germanischen Wortstellung. IF I: 333-436.

Wiedemann, Ferdinand J. 1875. Grammatik der Ehstnischen Sprache. L'Académie Impériale des sciences, St.-Péterbourg.

Zwicky, Arnold. 1977. On Clitics. Distributed by the Indiana University Linguistics Club.

Zwicky, Arnold. 1984. Particles and Clitics. OSU WPL 29: 105-116. (also appeared (1985) in Language 61.2: 283-305.

OSU WPL 34 (1986) 28-55

On Automatic and Simultaneous Syntactic Changes*

Brian D. Joseph

0. Introduction

Most linguists, in studying language change, have long assumed that there are changes which might well be described as being simultaneous, in that one change, D_X, occurs at the same time as another change, D_Y. In addition, it has also been assumed that there are changes which might be termed automatic, in that one change, D_X, necessarily causes another change, D_Y. In actuality, though, since the exact timing of changes is often hard to determine, it is generally the case that changes are counted as simultaneous if they at least appear to occur in close succession.

It should be clear that not all simultaneous changes are linked in the causal relationship implied by the label "automatic". In particular, two changes--for example a change in the articulation of some sound and a reanalysis of a syntactic construction--may have nothing to do with one another yet may just happen to occur (virtually) at the same stage in a language's development. More often, probably, two changes that are simultaneous--or nearly so, to be more accurate--do stand in a cause-effect relationship, so that one change can be taken to be a consequence of the other change. Even in such cases, though, there need not be any notion of necessity in the actuation of the second change, i.e. one does not have to be an automatic consequence of the other. Two examples from historical phonology demonstrate this difference well.

Martinet, in several works (e.g. Martinet 1953), has argued for the existence of so-called "drag-chains" in sound change, in which one sound shift leaves a gap in a system but "drags" another sound along with it to fill that gap. For example, under one possible interpretation of the Grimm's Law consonant shift in pre-Germanic, the shift of the Proto-Indo-European voiceless unaspirated stops, e.g. *t, to voiceless fricatives, e.g. *θ, left a gap in the consonant system that was then filled by the Indo-European voiced unaspirated stops shifting to voiceless unaspirated stops, e.g. *d ---> *t. In such an account, the *t ---> *θ change dragged along the *d ---> *t change. While Martinet has in general viewed such a second shift as a necessary consequence of the first, in actuality, sound systems tolerate many gaps happily, so the creation of such an imbalance in a system does not automatically occasion the filling of that gap through another sound shift. In such a case, then, two (virtually) simultaneous changes need not be causally linked.

An example of an automatic change, though, is provided by the restructuring of underlying lexical representations brought on by unconditioned sound changes. For instance, when Indo-European *d became Germanic *t, lexical forms which had had *d were restructured so as to reflect the new pronunciation, as in the change of the word for 'ten': /*dek'm/ ---> /*texum/ (cf. Gothic taihun, English ten, etc., and note that there were other changes as well not relevant here). At the point

at which *d became *t, there was no longer any support for underlying /d/ either from morphophonemic alternations or even distributional evidence, so the lexical form--under any set of theoretical assumptions about how such forms are established by speakers learning their language--would have to change when *d changed. Thus the restructuring would have been (virtually) simultaneous with the sound change and an automatic consequence of it.

This restructuring is an example of an automatic change from the realm of phonological change, but examples of automatic and simultaneous changes have been proposed for syntactic change as well. However, the cases that have been proposed for automatic and simultaneous syntactic changes are not without some problems. Accordingly, a brief review of some of these attempts at uncovering this type of syntactic change is undertaken here, and then two case-studies are presented from the history of Greek which provide stronger and more convincing instances of automatic syntactic changes.

It is important to point out, though, that it is very hard to prove conclusively that two changes are automatic or even that they are simultaneous; this is a recurring problem in the evaluation of such examples. However, where one can find either no evidence to the contrary or else positive indications that the two changes were not separated by long periods of time, it can be assumed that two changes which appear to be (virtually) simultaneous in fact are to be classified as such, for that assumption allows for the possibility of interesting claims regarding the nature of syntactic change.

Without the assumption of some kind of interaction between or among various changes, diachronic syntax becomes little more than a taxonomy of what changed between stage X and stage Y of a language; few, if any, interesting generalizations become possible about a theory of syntactic change, providing, for example, a delimitation of the range of possible changes in the syntax of a language. Therefore, wherever possible, the strongest position to take is that two apparently simultaneous changes are in fact simultaneous, for one can then work from there to try to find an explanation for this simultaneity. Accounting for one change in terms of another, by showing one to be an automatic consequence of the other, would be one way of providing such an explanation.[1]

Furthermore, such explanations for syntactic changes, when available, can be used as a way of constructing arguments for or against particular theoretical stances, under the assumption that a synchronic theory provides the constraints on possible changes a language may undergo.[2] Such a position has been taken, for example, by Lightfoot (1979a). He contends that Linguistic Theory should interact with a theory of change to pinpoint when grammars would undergo drastic reanalyses. One can further claim, following the line of argumentation being developed here, that a theory's ability to characterize one of two apparently simultaneous changes as being in fact an automatic change, a necessary consequence of and thus explained in terms of the other, should likewise count as an important criterion upon which to judge competing theories of grammar.

In particular, in the examination of the putative automatic changes from the history of Greek, an argument is developed against a derivational framework for a theory of syntax and in favor of a nonderivational approach. Briefly, a derivational theory of syntax is one in which rules apply in a certain order to produce a series of intermediate stages that convert a deep structure of a given sentence into a particular surface structure--the series of stages formed by the output of rule applications is called a derivation.[3] In a nonderivatinal framework, by contrast, there is basically no difference between deep structures and surface structures and thus one is not converted into the other via a series of intermediate steps; instead, some notion such as the designation of levels at which syntactic generalizations can be stated (e.g. initial syntactic level, final syntactic level, some combination, etc.)--as in current versions of Relational Grammar and Arc-Pair Grammar[4]--or some division of labor into components--e.g. semantic as opposed to syntactic, as in Generalized Phrase Structure Grammar with its rule-to-rule semantics--is employed. The analogue to a derivational theory's step-by-step rules in a nonderivational theory is a set of well-formedness conditions holding on surface forms, where elements can appear, in what combinations, how they relate to other elements in the sentence, and so forth. In such a system, the notion of derivation has no place.

Although comparing frameworks is a very tricky business, and perhaps not even possible, because the ground rules can be so very different in different frameworks, the two general approaches to syntax outlined here differ in one respect in the way they account for a particular change in Greek to be presented below. This distinction is discussed again in connection with that change after a look at some instances of automatic syntactic change that have been proposed in the literature.

1. Some Previous Attempts at Finding Automatic Syntactic Changes

Among the instances cited as examples of automatic syntactic change (though not necessarily labelled as such) are the following two provided by Lightfoot in various studies.

Lightfoot (1974, 1976, 1979a) has argued that a number of (nearly/virtually) simultaneous changes in the verbs which are now the Modern English modals (can, could, may, etc.) were the automatic consequences of a single innovative restructuring of the base rules in 16th century English. He claims that Old English and Middle English predecessors to the modals were real verbs, no different in any respect from other complement-taking verbs such as try or want, but that for a variety of reasons, they lost some verb-like features, e.g. no longer having full person and number paradigms, and were reanalyzed as forming a class distinct from that of try or want.

Thus, from a set of phrase structure rules as (1a) for Old and Middle English (which alternatively could have had the form in (1b)):

1. a. OE/ME: S ---> NP AUX VP
AUX ---> T(ense)

b. S ---> NP VP
VP ---> V + T (NP ...)

Lightfoot claims that 16th century English innovated a new expansion for AUX, creating a new deep structure category of modals (= M), and giving the rules in (2):

2. S ---> NP AUX VP
AUX ---> T (M)
M ---> can, may, ...

Furthermore, he claims that this new phrase structure rule, this single innovation, triggered the four changes listed below in (3) as automatic consequences:

3. a. no more infinitival forms of modals
b. no more gerund (-ing) forms of modals
c. only one modal per (simplex) sentence (in standard language)
d. no more have + MODAL + en combinations.

The absence of infinitival forms, for example, follows necessarily because the new modal class only occurred as a "sister" of T(ense) in the AUX node, a place where infinitives could not occur; similarly, the other changes in the modals are a necessary result of the nature of this new phrase structure rule.

In another work (1979b), Lightfoot proposes yet another instance of an automatic syntactic change, this time in the English passive, and this time the result of the addition to the grammar of a single transformational rule of NP Preposing. This rule led to the existence of a transformational rule of Passive whereas prior to the 16th century, Lightfoot claims, English had only a lexical passive rule. He is assuming a theory with a rigid distinction between lexical and transformational (or syntactic) rules, and argues that the properties of a syntactic rule as opposed to a lexical rule of Passive led to at least three automatic and simultaneous changes in English passives; in particular, three new passive sentence-patterns, listed below in (4), become possible:

4. a. passives with underlying indirect objects promoted (e.g. John was given a book)
b. "prepositional" passives (e.g. The terms were agreed upon)
c. passives with NP-subjects that do not bear a semantic relation to the main verb (e.g. John was expected to win).

These changes would have occurred automatically, Lightfoot claims, because the new transformational (syntactic) rule of Passive could move any NP after the verb to subject position, whereas, according to the properties of lexical rules the theory specifies, the lexical rule could only relate an active direct object with a passive subject. The sentence patterns illustrated in (4), then, could only arise with the advent of a transformational rule of Passive, so that these patterns are an automatic consequence, in Lightfoot's account, of the addition of such a rule to the grammar of English.

Despite these neat-looking examples of (virtually) simultaneous and automatic syntactic changes, Lightfoot's analyses are not uncontroversial and do not provide unqualified examples of this type of syntactic change. In particular, in each instance, one has to wonder whether Lightfoot has correctly identified cause and effect, i.e. is it really the case for the modals that the disappearance of the infinitival and gerund forms was the result of a base restructuring, or is it not possible that these forms were lost and only then was the grammar restructured to have a modal category to accomodate these now anomalous verbs? Furthermore, since other verbs do not have a full set of nonfinite forms--for some speakers, it seems that the verb stride does not have a past participle, with neither has strode, has stroden, has stridden, has strided, nor has stridded sounding acceptable[5]--the modals may just be a special case of the loss of nonfinite forms being generalized throughout a semantic class such as that formed by the modals.

Finally, exception can be taken to some aspects of Lightfoot's data. For example, regarding the modals, Lightfoot takes as significant the fact that the Oxford English Dictionary gives the last example of a modal infinitival or gerund form as occurring in the 16th century, even though the mere occurrence of a form in a text does not guarantee that it is still in current use (some texts are consciously archaizing, for example). Moreover, the process by which the verbs that ended up as the Modern English modals became specialized in their modal function and syntax was actually a very gradual change--for example, the gerund forms were rather rare at all prior stages of English[6]--and does not really display the suddenness that Lightfoot suggests. Thus, until the crucial examples that Lightfoot cites in support of his claims of simultaneity for these changes in question are subjected to careful philological scrutiny, his analysis has to remain tentative.

Moreover, Lieber (1979) has suggested that the factual basis for Lightfoot's claims about changes in English passive sentence patterns is faulty, for she finds in Old English passive sentences of the type Lightfoot says first appeared only in the 16th century. She concludes that Old English had a transformational (syntactic) rule of passive, as well as a lexical rule, and that the changes that in Lightfoot's account were simultaneous and automatic consequences of the addition of a rule of NP Preposing were features of Passive that were already present in the language. Such a finding, of course, if valid, renders this example of automatic syntactic change nothing more than a mirage.[7]

Besides these putative automatic, simultaneous changes due to restructuring or rule addition, there are also examples in the literature which invoke language universals, and claim that a change D_x automatically triggers another change D_y because D_x brings on a situation in which some universal is "activated", so to speak, and satisfying that universal requires the further change D_y. In such an instance, D_y is an automatic consequence of D_x and by the definitions adopted earlier, is simultaneous (or virtually so) with D_x. This type of explanation is evident in most of the work done recently on word-order change, in which putative universal correlations as in (5) have been called upon to explain, for example, a shift from postpositions to prepositions in Greek along with a shift in basic word order:[8]

5. OV with Postpositions / VO with Prepositions.

Another example involving universals is one described in Joseph (1978, 1980) concerning what happened to Greek Object Raising and Object Deletion sentences, i.e. constructions analogous to the English sentences in (6):[9]

6. a. Object Raising: John is easy to please.
   b. Object Deletion: Mary is pretty to look at.

when affected by the Greek replacement of infinitives by finite verbs.[10] In Greek, from Ancient Greek up to early Medieval Greek, there were Object Raising and Object Deletion sentences with a form entirely analogous to that found in English, with an infinitive in the complement clause and a zero-object (i.e. a missing object argument) with that infinitive on the surface.[11] Examples for pre-Modern Greek of Object Raising are given in (7) and of Object Deletion in (8), with a Ø marking the missing object argument:[12]

7. a. he: ergasia mathein Ø ... hraiste: edokei einai
      the-work/FEM.NOM learn/INF easiest/FEM.NOM seemed be/INF
      'The work seemed to be easiest to learn' (Xen. Oec. 6.9)
   b. tragoudousin to paranomon horo:sai Ø muste:rion
      sing/3PL the-illegal/NTR see/INF rite/NTR
      'They sing of the rite (which is) illegal to see' (Spanos 26 (12th century))

8. a. kai ... horan Ø stugnos e:n (Xen. Anab. 2.6.9)
      and see/INF gloomy/NOM was/3SG
      'And he was gloomy to look at'
   b. tous ... khrusinous hetoimous ekhei tou dounai Ø
      the-gold-pieces/ACC ready/ACC.PL has/3SG PRT give/INF
      'He has the gold pieces ready to give over' (Doukas 1164A, 13-14 (15th century)).

When the infinitive was replaced by a finite (i.e. person-marked and tensed) verb in late Medieval Greek, sentences corresponding to (7) and (8) continue in the language, but in a slightly different form; in particular, the late Medieval and Modern Greek continuation of the earlier Greek constructions now have a pronominal object in the complement clause that corresponds to the matrix subject. This is shown by the Modern Greek sentences in (9), where ta in (9a) and tin in (9b) are obligatory object pronouns in the complement clause:[13]

9. a. ta anglika$_i$ ine diskola na
      the-English/NTR.PL are difficult/NTR.PL PRT

      ta$_i$ katalavo
      them/NTR.PL understand/1SG

      'English is difficult for me to understand' (literally: "The English are difficult that I understand them")

b. i maria$_i$ ine omorfi na tin$_i$ kitazis
Mary/NOM is pretty/NOM.FEM PRT her/ACC look-at/2SG
'Mary is pretty to look at' (literally: "Mary is pretty that you look at her").

The change from a zero-object in the complement clause to an obligatory object pronoun can be taken as an automatic consequence of the change in the infinitive, i.e. in the nature of the complement clause verb itself, because of a putative universal constraint on Object Raising and Object Deletion constructions given in (10):[14]

10. Object Raising and Object Deletion cannot deprive a finite complement verb of its object.

This constraint is observed in several languages, including French, Spanish, German, Albanian, Irish, Korean, Mongolian, and Arabic (see Joseph (1978, 1980) for some discussion) and can account for the contrasts in (11) between unacceptable English Object Raising/Object Deletion sentences with finite complements versus acceptable ones with nonfinite complements:

11. a. *John would be difficult for me to imagine (that) I might invite Ø to my party
b. John would be difficult for me to imagine inviting Ø to my party
c. *Melina is too ugly for us to be able to convince John that he should kiss Ø
d. Melina is too ugly for us to be able to convince John to kiss Ø.

If this universal is valid, the change in the possibility of a zero-object as opposed to a pronominal object occurring in the complement clause of Object Raising and Object Deletion sentences would have been a necessary change, given the change in the type of complementation from nonfinite to finite.

A problem, though, with this account, and for that matter with any account making use of a universal, is that all universals are putative at best, subject to verification again and again as new data is brought to light, but therefore liable to be counter-exemplified by some piece of data not previously considered. For example, the potential Object Raising/Object Deletion universal given in (10) runs into some weak but nonetheless real counterexamples in English sentences such as (12):

12. a. ?A book like that is tough to claim that you've read Ø carefully.
b. ?This rock is too heavy to claim that I can pick up Ø.

While not fully acceptable sentences, nonetheless, those in (12) are not as bad as (1) predicts. Thus this universal has some validity, for the sentences in (12) are not wholly well-formed, but it cannot be an absolute universal. As a result, in a Medieval Greek Object Raising or Object Deletion sentence with a finite complement and a missing object with that finite verb, there would not have been any necessity for an object pronoun to arise, even if this might have been a likely or even

preferred development.

Similarly, many of the word order universals are of a statistical nature only, and meet with counterexamples (e.g. Papago and Persian appear to be exceptions to the correlations noted in (5)).[15] That being the case, one change would not necessarily force a second word-order change, at least as far as the correlation in (5) predicts.[16]

This cursory review of prior attempts to establish instances of automatic syntactic change shows that in order to get a good, i.e. relatively safe and unassailable, example of such a change, one needs cases in which either the data is clear or, if a universal is involved, it is one that is not controversial and can be supported by a wide range of relevant data. At this point, it is appropriate to examine two changes from the history of Greek which meet these requirements in order to demonstrate not only that automatic syntactic changes exist but also that they can be used in arguing for particular theoretical frameworks.

2. Copy-Raising in Greek

The two changes to be examined both involve and depend on a construction which can be referred to as "Copy-Raising". In order, then, to understand these changes properly, some background on this construction is needed. The Copy-Raising construction is one in which a nominal originating in--i.e. semantically linked (in initial structure) to--a complement clause appears superficially in a higher clause but shows an overt marker--in the form of a "copy" pronoun--of its presence in its "point of origin" (i.e. the lower clause). English sentences with the matrix predicate look like, as in (13a), have often been cited as examples of such a construction (the non-Raised version is given in (13b)):

13. a. Bill$_i$ looks like he$_i$ is ready to leave.
    b. It looks like Bill is ready to leave.

(13a), under such an analysis, would show a nominal (Bill) raised to subject status in the matrix clause with a copy (he$_i$) left in the complement clause.

This Copy-Raising construction is found in Classical and Hellenistic Greek, and has been studied in this context by Marlett 1976. Some examples are given in (14):

14. a. te:n ... huperbole:n to:n oreo:n ededoikesan
       the-pass/ACC the-mountains/GEN feared/3PL

       me: prokatale:phtheie (Xen. Anab. 3.5.18)
       lest be-occucpied/3SG

       'They were afraid that the mountain pass might be occupied' (literally: "They feared the mountain pass lest it be occupied")

b. epegino:skon de auton hoti houtos e:n
knew/3PL and him/ACC COMP this/NOM.MASC was/3SG

ho kathe:menos (Acts 3:10)
the-sitting/NOM.MASC.PPL

'And they recognized that he was the one sitting' (lit.: "They recognized him that he was the sitting one")

c. egno:n se hoti skle:ros anthro:pos ei
knew/1SG you/ACC COMP hard/NOM.SG man/NOM.SG are/2SG
'I knew that you were a hard man' (literally: "I knew you that you were a hard man") (Matt. 25:24)

Note that the copy is not always overtly present; because Greek has always been a language that suppresses unemphatic subject pronouns, a copy pronoun having subject status--as in (14a) and (14c)--does not have to appear on the surface. Marlett's analysis of these sentences as involving a Greek version of Copy-Raising, though admittedly not supported by hard evidence--such as demonstrably ungrammatical sentences and native speaker judgments--that is necessarily lacking for a "corpus language" such as Ancient or Hellenistic Greek, nonetheless can be adopted for two reasons. First, apparently synonymous non-Raising versions of such sentences can be shown to occur, and second, the logical structure of the predicates involved seems to point to a raising analysis (e.g. FEAR is a two-place predicate so that a sentence such as (14a) with a subject, a direct object, and a complement clause must not represent a basic structure with this verb).

Modern Greek also has a Copy Raising construction. Although it is not all that common a construction and is restricted to just a handful of verbs, nonetheless it is a construction-type in the language and so must be part of any descriptively adequate grammar that might be constructed for the language.[17] An example of this construction is given in (15):

15. θeoro tin maria pos mono afti ine eksipni
consider/1SG Mary/ACC COMP only she/NOM is/3SG smart/NOM.FEM
'I consider only Mary to be smart' (literally: "I consider Mary that only she is smart")

and it can be contrasted with a synonymous non-Raised version as in (16):

16. θeoro pos mono i maria ine eksipni
consider/1SG COMP only Mary/NOM is/3SG smart/FEM
'I consider only Mary to be smart' (literally: "I consider that only Mary is smart").

For Modern Greek Copy Raising sentences, the usual range of arguments based on selectional restrictions, idiom chunks, active/passive synonymy, Reflexivization, and Passivization are all available to show that the accusative noun phrase in such sentences corresponding to tin maria in (15) is in fact a main clause direct object on the surface and moreover corresponds to an initial-structure complement clause subject.[18] Most commonly in Modern Greek, as in earlier stages of the language, these Copy Raising sentences have no overt copy pronoun in the complement clause, for unemphatic subject pronouns generally do not appear on the

surface in Modern Greek; the occurrence of mono 'only' in (15) provides the emphasis necessary for the occurrence of the pronominal copy afti.

With these preliminaries concerning the nature of the Copy Raising construction out of the way, a description and account of the changes in this construction between (early) Post-Classical Greek and Modern Greek can be developed. The first change concerns a change in the form, specifically the person, of the complement clause copy pronoun in one subclass of Copy Raising sentences.

3. The Change in Complement Clause Agreement

In Cópy Raising sentences in earlier stages of Greek, the raised nominal and the copy always were of the same grammatical person, as shown by the examples given earlier; for example, in (14b), both auton, the raised nominal, and houtos, the copy pronoun, are third person. Moreover, in sentences in which the raised nominal was coreferent with the matrix clause and therefore of the same person as the matrix subject, it appeared on the surface as the reflexive pronominal form; in such a sentence, then, the complement-clause copy likewise was of the same person as the matrix clause subject (and object). An example of such a Copy Raising cum Reflexivization sentence is given in (17), where the main clause subject--understood to be the second person plural form humeis (absent on the surface because it is unemphatic in this sentence)--the reflexive form heautous and the raising copy humin are all second person forms:

17. ouk epigino:skete heautous hoti ie:sous
not know/2PL yourselves/ACC COMP Jesus/NOM

khristos en humin (2Cor. 13:5)
Christ/NOM in you/DAT.PL

'Do you not understand that Jesus Christ is within you?'
(literally: "Do you not understand yourselves that Jesus Christ (is) within you?")

In Modern Greek Copy Raising sentences with Reflexiviation in the matrix clause, though, a difference from the earlier pattern exemplified in (17) is evident. In particular, the raised nominal (the reflexive form) and the complement clause copy still agree in person (as in (14) and in (17)), but they are both different in person from the matrix subject. Such a sentence is given in (18), where the matrix clause subject is first person (suppressed as unemphatic, but understood to be ego 'I/NOM'), while the reflexive form (ton eafton mu) is third person and the raising copy (aftos--in parentheses since it is overt only if emphatic) is also third person:

18. den θa afiso ton eafto mu na petixi (aftos)
not FUT let/1SG the-self/ACC my PRT succeed/3SG he/NOM
'I won't let myself succeed' (literally: "I won't let the self of me that he succeed").

Even though, as noted above, the raising copy can be absent on the surface if, as is most usually the case, it is unemphatic, the third

person verb agreement in the complement clause gives an indication of what person its subject is understood to be, and shows that it is different from that found in the matrix clause (as shown by the matrix clause first person agreement). There is thus a syntactic difference in Copy Raising sentences between Hellenistic Greek and Modern Greek in the person of the raising copy (and thus in the verb agreement in the complement clause when the copy is the subject) when the raised nominal is subject to Reflexivization in the main clause. It is this syntactic difference which demands an explanation.

One fact which is relevant for an explanation of this change is that the reflexive form itself has a different character in Modern Greek from that found in earlier stages of the language. In particular, in Ancient Greek and Hellenistic Greek, the reflexive was a pronominal form, agreeing in person with its antecedent and showing different forms for the different persons; a portion (accusative case only) of the paradigm for the Ancient Greek reflexive pronoun is given in (19):

19.	1SG	emauton (MASC)	emaute:n (FEM)	
	2SG	seauton	seaute:n	
	3SG	heauton	heaute:n	heauto (NTR)
	1PL	he:mas autous	he:mas autas	
	2PL	humas autous	humas autas	
	3PL	heautous	heautas	heauta (NTR)

and the Hellenistic paradigm was similar.[19] In Modern Greek, though, the reflexive is itself a fixed form, consisting of ton eafto, literally "the self", with a possessive pronoun; it thus has the form of a possessed nominal,[20] being literally "the self" with a possessive pronoun and thus structurally parallel to ton adelfo 'brother' with a possessive pronoun. The reflexive form, therefore, is syntactically a third person nominal, just as is any nonpronominal noun phrase. The possessive pronoun is the only thing in the reflexive form that necessarily[21] shows agreement with the reflexive antecedent. A partial (accusative only) paradigm of the reflexive form is given in (20), along with the possessed form of 'brother' for comparison:

20.	1SG	ton eafto mu	1PL	ton eafto mas
	2SG	ton eafto su	2PL	ton eafto sas
	3SG	ton eafto tu (M/N) ton eafto tis (F)	3PL	ton eafto tus (M/N/F)
	cf.	ton adelfo mu/su/tu/tis/mas/sas/tus 'my/your/his, its/her/our/your/their brother'.		

This change in the form of the reflexive took place at least as early as the 12th century; an example from the poems of Glykas is given in (21):

21. na pnikso: ton heauton mou (Glykas 288 (12th cent.))
FUT drown/1SG the-self/ACC my
'I will drown myself'.

There has thus actually been a multiple change in the form of Raising cum Reflexive sentences between earlier stages of Greek and Modern Greek. The reflexive form and the copy found in Raising cum

Reflexive sentences have both changed so as to differ from the matrix subject (the reflexive antecedent) in person. It seems, therefore, that two (virtually) simultaneous changes have occurred, and moreover, it can be shown that this set of changes provides an excellent candidate for consideration as an automatic syntactic change. In particular, it seems that the change in the form of the complement clause copy pronoun in such raising sentences is an automatic consequence of the change in the form of the reflexive.

The motivation behind the necessary change of the copy to third person along with the change of the reflexive to third person status is the following universal:

22. A copy must agree in all relevant features (i.e. relevant to the language in question, e.g. person and number) with the nominal of which it is a copy.

It is safe to say that most linguists would agree that (22) is a fairly uncontroversial universal, one for which it would be very hard to find counterexamples. For instance, one class of apparent counterexamples, namely sentences such as (23):

23. John, I can't stand the idiot

is probably best analyzed as involving dangling topics rather than a nonagreeing copy. Furthermore, the need for such a universal independent of the Greek Raising cum Reflexive sentences under consideration is shown by copy/antecedent agreement in a variety of constructions in a variety of languages, including English look-copy sentences (cf. (13) above) and Left-Dislocation sentences, as in (24):

24. a. The scissors$_i$ look like they$_i$/*it need to be sharpened.
    b. John, I can't stand him/*her/*it.

If this universal is valid, then a ready explanatin is provided for the apparent automatic nature of the syntactic change in question here. When the reflexive changed in person, the universal would guarantee that the copy pronoun in the complement clause--inasmuch as it is a copy of the nominal that surfaces as the reflexive--would also change.

The only potential problem with this account--and thus with taking this change as a true instance of an automatic syntactic change triggered by a universal--is the fact that it cannot be conclusively demonstrated that the change in the complement clause copy was simultaneous with the change in the reflexive. Unfortunately, as is so often the case in such investigations, there is not very much historical data to back up the claim. However, in this instance, there is at least some. The reflexive change, as noted above (cf. (21)), took place no later than the 12th century. The first Raising cum Reflexive sentence, though, that occurs in Greek after the reflexive change dates from the late 16th century, from the Cretan comedy Katzourmbos:

25. thelo: aphe:sei ton emauto mou ... na ...
want/1SG let/INF the-self/ACC my PRT

ksale:smone:se: (ton nou mou) (Katz. II.173-4)
forget/3SG the-mind/ACC my

'I won't let myself forget my mind'.

In interpreting this sentence as a relevant example for the discussion here, it is necessary to take the third person singular verb agreement exhibited by the complement verb ksale:smone:se: as indicating that the complement clause subject--absent on the surface because of unemphatic subject pronoun drop in Greek--was third person. What makes this sentence valuable, despite the four centuries between it and the reflexive change is the fact that no counterexamples, at least, with agreement patterns different from those in (25), are to be found. Sentences such as these, by their very nature, might well be expected to be quite rare,[22] so that despite the scarcity of conclusive historical data, there is really nothing standing in the way of taking (25) and its kind to represent a change that occurred concomitantly with the changes in the reflexive evidenced in (21). This consideration and the strong--and generally unassailable--universal in (22) that has been proposed to be at work in the change combine to make this a very good candidate for a real example of an automatic syntactic change.

4. Sneaky Passives Diachronically

Another aspect of the Copy Raising construction provides another instance of a syntactic change which, though lacking in some of the crucial historical data, nonetheless seems to be a real instance of an automatic syntactic change. In this case, however, there are some interesting theoretical dividends concerning differences between derivational versus nonderivational frameworks that can be reaped from the account of the change.

This change concerns the status of Copy Raising sentences in which the complement clause is passive and the raised nominal corresponds to the agentive noun phrase in the complement clause. An example of such a sentence from Ancient Greek is given in (26):

26. dedoik' emauton ... me: poll' agan
fear/1SG myself/ACC not much/NOM.PL.NTR too

eire:mena e: moi (Oed. Tyr. 767)
said/PASS.PPL.NOM.PL.NTR be/3SG me/DAT

'I fear that too much has been said by me'.

Such sentences can be referred to as "Sneaky Passives", following Perlmutter & Soames (1979: 164ff.). because in a derivational framework, these can be derived by applying Copy Raising in the matrix clause and then applying Passive "sneakily" into the complement clause; this latter step is possible because Copy Raising leaves a fully-intact complement clause, complete with subject (the copy pronoun) and object, and thus meeting the structural requirements for application of a passive rule.

This derivation is sketched in (27):

27. UNDERLYING STRUCTURE: $_S$[I fear $_S$[I say too much]$_S$]$_S$

COPY RAISING : [I fear myself [I say too much]]
(+ REFLEXIVE)

PASSIVE ("SNEAKILY"): [I fear myself [too much be said by me]]

where many details of structure have deliberately been left out.

Sneaky Passives were fully grammatical in earlier stages of Greek, as (26) indicates. It is important to note that Copy Raising could in general operate on oblique nominals, as shown by examples (17) above and (28) below:

28. a. phoboumai humas me: eike: kekopiaka eis humas
fear/1SG you/ACC.PL not in-vain worked/1SG among you/ACC.PL
'I fear that in vain have I spent my labor among you'
(literally: "I fear you lest I have spent my labor among you in vain") (Gal. 4:11)

b. kai poiei pantas$_i$ ... hina do:sin autois$_i$ kharagma
and cause/3SG all/ACC.PL COMP give/3PL them/DAT mark/ACC
'And he caused everyone ... to receive a mark'
(literally: "He caused everyone$_i$ that they$_j$ give to them$_i$ a mark") (Rev. 13:16)

The fact that oblique nominals were eligible for Copy Raising means that a "Sneaky Passive" sentence such as (26) actually has two possible derivations. Besides the one outlined in (27), there is also a derivation in which passive applies in the complement clause followed by Copy Raising of the agent nominal created by passive into the higher clause. It is significant that (28b) shows Copy Raising of a dative noun phrase (autois), since it shows that even if the eligibility conditions for Copy Raising were stated in terms of case-marking, such a derivation for (26)--where the nominal corresponding to the raised noun phrase in the complement clause is a dative, moi--cannot be ruled out. It can be concluded, then, that a sentence such as (26) indeed has two possible derivations--the "Sneaky" Passive derivation and the one in which first Passive applies in the lower clause and then Copy Raising in the higher clause.

Sometime between Ancient Greek and Modern Greek, and unfortunately there is absolutely no textual evidence to indicate when this may have occurred, a change took place in Copy Raising. Whereas in Ancient Greek, it seems that any nominal could be raised to object status in a higher clause (cf. the raising of obliques in (28)), in Modern Greek Copy Raising is restricted to operating only on subjects.[23] Thus, sentences such as (29) are ungrammatical:

29. *θeoro ton yani pos ton vrika ili0io
consider/1SG John/ACC COMP him/ACC found/1SG stupid/ACC
'I consider John that I found him (to be) stupid'.

Since there is no indication regarding the status of such sentences in Medieval Greek, the assumption can be made that the Modern Greek evidence shows that the change in which nominals can be raised via Copy Raising has taken place within the Modern Greek period, i.e. in what is roughly the contemporary language.

What is interesting here is that besides this change in the eligibility of nominals for Copy Raising, there is another change, namely a change in the status of Sneaky Passive sentences. In Modern Greek, Sneaky Passive sentences are ungrammatical, as shown by (30):

30. *Θeoro ton yani pos i maria vlaftike apo afton
consider/1SG John/ACC COMP Mary/NOM hurt/3SG.PASS by him/ACC
'*I consider John that Mary was hurt by him'.

Thus it appears that this is another instance of a simultaneous change--a change in what can raise occurs together with a change in the status of Sneaky Passives. Although the mere apparent (or stipulated) simultaneity of the two changes does not mean that one is an automatic consequence of the other, as noted above in section 1, it is desirable to treat them as having that relationship, for then one change can be explained in terms of the other. Thus, following that line of reasoning, an account is developed below in which the change in Sneaky Passives is an automatic consequence of the change in the Copy Raising construction; any such account necessarily is stronger--and thus more interesting and preferable--to one in which the changes are unrelated (again, as noted in section 1).

In a nonderivational framework, as pointed out in section 1, there are no syntactic rules that work to convert a deep structure into a surface structure via a series of phrase-markers (or the equivalent). Instead, sentences are generated in their surface forms in accordance with the analogue of syntactic (transformational) rules, namely well-formedness conditions on these surface strings. The analogue of Subject-to-Object Raising, for instance, would be a well-formedness condition sanctioning the occurrence of a nominal that is semantically "relevant" (to use as theory-neutral a term as possible) only in a lower clause (i.e. it bears a logical relation only in that clause) as an object in a higher clause. Put in terms of a nonderivational framework with a recognition of syntactic levels and grammatical relations, such as Arc Pair Grammar or Relational Grammar, such a condition would allow an initial level complement clause nominal, e.g. a subject, to occur as a final level matrix clause object.

In such a framework, Ancient Greek Copy Raising would be a well-formedness condition such that a nominal bearing any final level grammatical relation in the complement clause may "legally" be the matrix clause object. This condition is stated in slightly more formal terms in (31):

31. Condition on Ancient Greek Copy Raising:

The final complement GR_x (= any grammatical relation)
is the final matrix clause GR_2 (= direct object).

In other words, (31) allows the occurrence of Copy Raising sentences in which the final matrix object bears any final level grammatical relation in the complement clause.

A condition such as this sanctions Sneaky Passives such as (26), repeated here for convenience:

26. dedoik' emauton ... me: poll' agan
fear/1SG myself/ACC not much/NOM.PL.NTR too

eire:mena e: moi (Oed. Tyr. 767)
said/PASS.PPL.NOM.PL.NTR be/3SG me/DAT

'I fear that too much has been said by me'.

because the matrix object emauton--which does not bear a logical relation to the matrix verb--bears some final grammatical relation in the complement clause, namely the relation of the agent in a passive clause (the "Chômeur" relation of Relational Grammar, the 8-relation of Postal's Arc Pair Grammar). The well-formedness condition states only that this nominal must bear some relation in the lower clause; it does not restrict which relation this might be, so that passive agent meets the requirements of the condition.

In Modern Greek, though, the well-formedness condition for Copy Raising sentences has changed so that instead of being able to be any complement clause grammatical relation, the matrix object can only be the final complement clause subject (as noted earlier--recall example (29)). The Modern Greek version of this condition is given in (32):

32. Condition on Modern Greek Copy Raising:

The final complement GR_1 (= subject) is the final matrix clause GR_2 (= direct object).

This differs from (31) just in the specification of GR_1 as opposed to GR_X, that is, a relatively minor change from a formal standpoint. However, it is a change that has important consequences. In particular, (32) automatically rules out Sneaky Passives because in Sneaky Passives, the matrix object in a Copy Raising is not the final complement clause subject, but rather is the final passive agent (i.e. chômeur or 8-relation).

There are admittedly a few potential problems with this account. In particular, since in Greek Raising there is a copy of the raised nominal in the lower clause, is it the copy or the matrix object that is considered to bear the relevant grammatical relation in the lower clause? It may be necessary to stipulate something to the effect that a copy counts the same as the form of which it is a copy in terms of satisfying the well-formedness conditions and the "is" relationship utilized in the Raising well-formedness conditions (i.e. in (31) and (32)). Alternatively, the existence of a chain of "control" linking the matrix object with its copy in the lower clause may be sufficient. This particular problem, however, is not a problem just for Copy Raising but

rather is a general problem of determining how the overall syntactic framework should treat copies of nominals that bear particular grammatical relations.

Still, in this framework, the problem posed by the apparent simultaneity of these two changes--the change in what can raise and the change in the status of Sneaky Passives--finds a straightforward solution, for the change from (31) to (32) automatically triggers the change in the acceptability of Sneaky Passives. The ease with which the nonderivational framework can account for these two changes is striking, for it contrasts with the extra machinery and extra assumptions needed in a derivational account of the changes.

For convenience in exposition, a derivational framework is assumed here in which rules are stated in terms of grammatical relations; this decision allows for easier comparison with the nondeirvational account outlined above, for rules in that account are so stated. The main assumption behind a derivational approach is that syntactic rules convert an initial structure into a surface structure through a series of sequentially-ordered steps (phrase markers, in the terminology of transformational grammar). Raising, therefore, in such a framework is a "process" by which a nominal in a lower clause becomes the object in a higher clause. For the sentence-type under consideration, i.e. sentences parallel in structure to (26), it has been established (see above, earlier in this section) that two derivations--two sets of sequentially applied syntactic rules--are possible: first Passive in the lower clause followed by Raising in the higher clause but also the "Sneaky" Passive derivation with first Raising in the higher clause and then Passive applying "sneakily" in the lower clause left intact by Copy Raising.

It is well-known that the assumption of sequentially-ordered rules in syntax has led to the recognition of the need for the cyclic application of syntactic rules. That is to say, natural languages exhibit syntactic phenomena, well-discussed in the literature,[24] which require recourse to a device such as cyclic rule application in order to be accounted for in a derivational framework. In cyclic rule application, syntactic rules apply as a block sentence by sentence from the most deeply embedded clause in the phrase marker to the topmost (matrix) clause. In a derivational framework, therefore, the cycle has been posited as a linguistic universal.

Similarly, along with the cycle, it turns out that there are phenomena in natural languages which require the imposition of a constraint--the Strict Cyclicity constraint--which prevents a rule from applying (or reapplying, as the case may be) into an already cycled-on domain. With such a constraint, once a higher clause has been reached in the cyclic application of rules, a lower clause--an already cycled-on domain--would not be a possible domain for a rule. With Strict Cyclicity, rules cannot "reach down", so to speak, so as to apply entirely within a cyclic domain that has already been passed. As with the cycle itself, the Strict Cyclicity condition has been proposed as a linguistic universal.

However, in at least some versions of derivational frameworks, there are rules which can be called noncyclic or postcyclic, i.e. they

are not "in" the cycle. These rules, moreover, can apply freely into already cycled-on embedded clauses. An example of such a rule would be Relativization or Question Movement;[25] thus a question word which originates in an embedded clause nonetheless can be fronted when the matrix clause is reached after the cycle, as in (23), where the Ø indicates the deep structure point of origination of the question word:

33. Who did Sally think John felt Bill was ready to hit Ø?

One interpretation of this observation is that the principle of Strict Cyclicity is valid only for cyclic rules, and does not hold, universally it would be posited, for rules not in the cycle (whether demonstrably post-cyclic or simply not demonstrably cyclic) rules. This is an interpretation which becomes crucial later on in the discussion.

Thus, a derivational framework has derivations, it has the cycle, and it has a principle of Strict Cyclicity that is restricted to cyclic rules. Given these elements of the framework, the change in Greek Copy Raising would automatically trigger a change in the status of Sneaky Passives. For Ancient Greek, the framework just outlined would allow only the derivation of Raising cum Passive sentences such as (26) only via the derivation in which Passive applied in the lower clause and then Raising applied in the higher clause to raise the now passive agent to object status (recall that in Ancient Greek, Raising could operate on nonsubject nominals). The other possible derivation--the "Sneaky" Passive derivation by which first Raising applied in the higher clause and then Passive "snuck" down into the lower clause to apply and put the original subject into an agentive phrase--would be ruled out because it would violate the principle of Strict Cyclicity (by reaching down into an already cycled-on domain).

Therefore, the change in which nominals were eligible for Raising (see (31) and (32) above) would be reflected also in a change in the status of Sneaky Passive sentences such as (30), because the only way (30) could be derived in this framework is by Raising an oblique (agentive) nominal (the other derivation being ruled out by Strict Cyclicity). A restriction on what can raise--from any nominal to only subjects--therefore automatically leads to a situation in which the only possible derivation for a sentence is systematically ruled out. Such a sentence is thus ungrammatical, for it cannot be derived.

At this point, from a comparison of the derivational account just presented and the nonderivational account preceding it, it would appear that the two accounts are equivalent. In both accounts, the change in Sneaky Passives falls out as an automatic consequence of the change in Copy Raising acting in concert with certain aspects of each framework that are either built-in or are universal parts of the theory in question (as the cycle is in a derivational framework).

However, on closer inspection, it turns out that there is a crucial difference between the two accounts. In particular, the derivational account must make one further, unwarranted and thus unmotivated assumption.

It came out earlier in the discussion that under certain interpretations of the way in which rules such as Question Movement operate, the principle of Strict Cyclicity would have to be valid only for cyclic rules. The consequence of such a restriction of this principle, however, is that only as long as Passive is a cyclic rule will it be constrained by Strict Cyclicity so as not to apply "sneakily" into an already cycled-on domain. That means that the derivational account must make the additional assumption that the rule of Passive stayed as a cyclic rule between Ancient Greek and Modern Greek, i.e. that it maintained its "cycle-type" and did not become noncyclic. This is a result which could be guaranteed by an appropriate typology of cyclic rules, such that a rule with properties such as the Passive rule has would necessarily be a cyclic rule, but in the absence of such a typology,[26] it would constitute an extra assumption necessary in a derivational framework. This result would mean further that the ungrammaticality of Sneaky Passive sentences in Modern Greek really is not an automatic consequence of the change in Copy Raising, for Sneaky Passives could have remained grammatical if Passive had changed its cycle type and become a noncyclic rule (and thus not subject to Strict Cyclicity).

Furthermore, there is some evidence to suggest that rules can change their cycle type diachronically and move from being a cyclic rule to being noncyclic. In particular, Reflexivization (at least the so-called "Direct Reflexivization") in Ancient Greek, as best as can be determined, was probably a cyclic rule, inasmuch as it interacts with apparent EQUI and Raising constructions much as Reflexivization in English interacts with the English analogues of those constructions. Since that interaction in English has generally been taken as evidence for the cyclicity of Reflexivization in English, a similar conclusion can be drawn for Ancient Greek Reflexivization.[27] However, in Modern Greek, due in part to the change in the morphological make-up of the reflexive expression (see the discussion in section 3 regarding (19) and (20)), Reflexivization must be taken as a noncyclic rule. The crucial sentences that lead to this conclusion are the following:

34. a. den Θa afiso ton eafto mu na me katastrepsi
NEG FUT let/1SG the-self/ACC my PRT me/ACC destroy/3SG
'I won't let myself destroy myself' (literally: "I won't let myself that he destroy me")

b. *den Θa afiso ton eafto mu na katastrepsi ton eafto tu
his

c. *den Θa afiso ton eafto mu na katastrepsi ton eafto mu
my.

If Reflexivization were cyclic, then it would be expected that (34c), with multiple occurrences of ton eafto mu, in both the lower clause and the higher clause, would be grammatical. However, (34c) is ungrammatical, as is (34b), where there is a reflexive form in the lower clause but the possessive that occurs with it is third person (agreeing with the person of the reflexive in the higher clause). The only acceptable version of such a sentence with Raising and Reflexivization[28] is that given in (34a), in which there is Raising and Reflexivization in the higher clause but no evidence of Reflexivization in the lower clause. Such facts run counter to the predictions made by an assumption

of cyclicity for Reflexivization in Modern Greek, so it can be concluded that the rule is noncyclic in this stage of the language and that therefore Reflexivization has changed its cycle-type between Ancient Greek and Modern Greek.[29]

An additional example of a change in cycle-type of a rule has been proposed by Haiman 1974. He argues that various phenomena connected with the Verb-Second constraint in Germanic languages first entered particular grammars as postcyclic processes, subsequently became cyclic, and in some cases forced a deep structure reanalysis. For instance, he points to the fact that many dummy pronoun insertions (e.g. the it of English it-Extraposition, there of there-insertion, etc.) are cyclic in modern Germanic languages but appear to be postcyclic in earlier stages of the langauges in question. A similar analysis is offered for the Verb-Second Constraint itself.

These two examples, Greek Reflexivization and the Germanic Verb-Second constraint, suggest that a change in cycle-type is a possible type of change that a rule (or grammatical constraint) can undergo in a derivational theory. That being the case, the change in Sneaky Passives can only be an automatic consequence of the change in what can undergo Raising if it is assumed that Passive stayed a cyclic rule in Greek (or if an adequate typology of cyclic rules is developed--see footnote 26); since there is no reason, in the absence of a suitable typology, why Passive should remain cyclic, it must be concluded that the derivational account cannot adequately characterize a relationship between the change in Sneaky Passives and the change in Copy Raising.

Thus the nonderivational account actually provides a better explanation of these changes in Greek than the derivational account does, for it does not require the additional ad hoc assumption regarding the maintenance of cycle-type for Passive. Consequently, the nonderivational account is to be preferred. These differences are summarized in (35):

35. In order for the change in Sneaky Passive to be automatically accounted for:

    a. a derivational account needs:

        i. the cycle and a principle of Strict Cyclicity valid for cyclic rules (this is given by the theory)
        ii. the change in the Raising rule (see (31) and (32))
        iii. the assumption that Passive maintains its cycle-type and is therefore subject to Strict Cyclicity

    b. a nonderivational account needs:

        i. a notion of levels to which syntactic rules can make reference (this is given by the theory--in the version of nonderivational grammar assumed here)
        ii. the change in the Raising rule (see (31) and (32)).

The nonderivational account has no derivations and therefore no cycle; the problem of change in cycle-type is, for such a framework, really only a pseudo-problem, for it is one that is forced only by the

ground rules of the derivational framework. Thus this syntactic change in Greek makes it clear how a derivational framework is burdened by all sorts of extra devices and machinery, such as the cycle, as well as problems, such as a change in cycle-type, that are, in a real sense, nothing more than artifacts of these extra devices.

5. Conclusion

By way of conclusion, a summary of the results developed here can be given, as well as a recognition of some remaining problems.

First, it should be clear that automatic syntactic changes do exist, even though some of the previous attempts at uncovering such changes are probably not as conclusive as they might appear at first to be. The examples involving changes in the Greek Raising construction seems to be fairly good cases of automatic syntactic change.

Second, it is also clear that different theoretical frameworks handle the same syntactic change in different ways, demonstrating that the view that emerges of what changes there are in a language is to a large extent colored by one's view of how synchronic grammars are to be characterized. It is possible, moreover, to compare the accounts that are provided in different frameworks. Lightfoot (1979a) has claimed that diachronic syntax provides "a new style of argumentation for choosing between competing theories and synchronic descriptions, by requiring that the theory interact with a theory of change to account for the point at which grammars undergo reanalyses or 'catastrophic' changes". The suggestion here is that the ability to relate simultaneous changes as being automatic changes is another criterion for deciding between competing theories to which diachronic syntax can contribute.

Third, given such a criterion for deciding between competing theories, the evidence from the change in the status of Sneaky Passive sentences in Greek shows that a nonderivational approach to syntax is to be preferred over a derivational approach, for the latter provides a simpler and less ad hoc account of the Sneaky Passive change and of the means for connecting the Sneaky Passive change with the Copy Raising change in a manner that is both natural and automatic.

Finally, it is only fair to mention what seems to be the only real problem with the discussion of these changes in Greek syntax, namely the lack of conclusive historical evidence. There are clear indications about the status of the relevant constructions in the first stage, Ancient Greek (i.e., in this context, Classical and New Testament Greek), and about the last stage, Modern Greek, but only meager indications about the intervening stages, especially in the case of the Sneaky Passive construction. Unfortunately, there is nothing that can be done about the lack of data;[30] it is simply a fact of life in historical studies that crucial data can be missing. One might simply dismiss these otherwise very interesting examples for that reason or else take them at face value and try to grapple with them. The latter approach has been adopted here, for the former seems counter-productive in that it limits the extent to which an understanding of syntactic change can be increased.

Notes

*This paper was read as an invited lecture to the Indiana University Linguistics Club in March of 1980. What is published here is essentially that version, with a few relatively minor changes in organization and diction, as well as updating of references, providing of footnotes, etc.

1. There are of course other modes of explanation that one might explore, such as a functional or social motivation for the occurrence of one change or another. It is my belief that no one type of explanation, whether formal (such as is pursued here), functional, social, or whatever, precludes the investigation of causation from one of the other perspectives, i.e. there is not only a formal side to language (how it is represented in a speaker's individual competence), but also a functional side (the uses to which a speaker puts this competence), a social side, etc.

2. This assumption is justified by the view in which diachrony is seen as the transition of a language through a succession of synchronic stages. Linguistic theory provides the constraints on possible synchronic systems, so that the movement through time is always the movement of one possible system into another possible system. Therefore, a theory of synchrony must be able to account for (at least certain aspects of) diachrony as well. While it may be true that not all linguists would adhere to such a view, it is not an idiosyncratic one; Culicover (1984: 118), for example, labels such an assumption "reasonable".

3. Although the issue of whether grammars are derivational or not is less an issue now than when this paper was first written (1980), it is important to note that some still-current theories are derivational in principle, including Government and Binding Theory (despite the considerable limitation on the number of rules that apply to convert deep structures into surface structures), while others, including Generalized Phrase Structure Grammar, are distinctly nonderivational.

4. See, for example, Postal 1986, and the papers in Perlmutter 1983 and in Perlmutter and Rosen 1984.

5. To my ear, <u>has stridden</u> does not sound all that bad. As Jeffrey Huntsman of Indiana University has kindly pointed out to me, the form <u>stridden</u> does occur in a variety of texts and there is also the form <u>bistridan</u> in Old and Middle English; moreover, the oddity of a perfect tense formation with <u>stride</u> may be more a function of the relative infrequency of the base verb itself, and not a peculiarity associated with the past participle.

6. I am grateful once more to Jeffrey Huntsman of Indiana University for reminding me of this important fact.

7. The debate perhaps should not be closed on this point, however, for Russom (1982) has shown that Lieber's examples of syntactic passives in Old English have been misanalyzed. I leave this matter to future research, inasmuch as my intent here was only to provide an example that

has been given in the extant literature of a change that would be simultaneous and automatic in the sense developed here.

8. The literature on word-order change and the relevance of universals is enormous, and no attempt can be made here to even briefly survey the field; the reader is referred to Hawkins (1979, 1980, 1982a, 1982b, 1983), Smith 1981, Coopmans 1984, and Payne 1985 for discussion.

9. The exact analysis of these sentences is irrelevant to the point being developed here. In particular, in some theoretical frameworks, e.g. Government & Binding Theory, there is no "raising" at all in (6a) but rather WH-Movement, and in nonderivational frameworks, there are no raising or deletion processes at all. Moreover, in some accounts, both sentence-patterns in (6)--despite some differences they exhibit, for example, in the possibility of extraposition--are treated as involving the same type of syntactic structure. Such considerations are irrelevant to the matter at hand because in any theoretical framework, sentences such as those in (6) and the construction-type(s) they represent must be accounted for in some way, and the point regarding automatic changes in the Greek equivalents of these patterns can thus be translated into the appropriate theoretical analogue.

10. For details on the loss of the infinitive in Greek, see now Joseph 1983 and references therein.

11. There is of course also a missing subject argument as well with the infinitive, a fact which is less significant for the changes to be described here than the missing object.

12. For further examples and discussion, see Joseph (1978, 1980, 1983).

13. It is important to point out that the complement object was obligatorily missing in these constructions from Ancient Greek up through early Medieval Greek. Examples of these constructions from Medieval Greek are to be found in Joseph (1978, 1980, 1983).

14. I purposely state the constraint in this form, mentioning specific constructions, and not in some more general manner (e.g. utilizing the various versions of the Tensed S Condition of Chomsky 1973) in part because it is irrelevant just how generally this constraint holds in the grammar, as long as it is valid for the constructions under consideration, and in part because I feel that there is sufficient reason not to believe that the Tensed-S condition is a valid condition (cf., for example, the possibility of WH-Movement out of tensed clauses--though such constructions admittedly have a different analysis in Chomsky's framework).

15. Hawkins (op. cit.), though, does attempt to address this problem through the use of implications involving more than two elements, which he claims leads to better statistical validity (e.g. SOV ---> (ADJ + N ---> GEN + N)).

16. This point has been made most recently by Payne 1985, though see also Coopmans 1984.

17. The exact analysis of this construction has been the subject of considerable debate among generative grammarians of Modern Greek; see, for example, Joseph (1976, 1978: Chapter 10, 1980), Philippaki-Warburton (1979, 1986), Kakouriotis 1979, Ingria 1981, inter alia. The facts under consideration remain the same, and the debate has been spurred in part by changes that have been made in the assumptions that underlie the theoretical frameworks adopted by different linguists.

18. See the references in footnote 17, especially Joseph (1976), for details. Note that even in those frameworks (e.g. that of Ingria 1981 or Philippaki-Warburton 1986) in which no raising is assumed for this construction, there is obligatorily a "linkage" between the matrix accusative nominal and the complement clause subject position. Thus, the account to be given here of changes in this construction, an account which, given the framework adopted here, depends on the assumption of Copy-Raising, could be translated into a framework with no raising through the use of this obligatory linkage.

19. Admittedly, the second person plural form in (17), heautous, differs from the Ancient Greek form, and shows generalization of the older third person plural form; however, heautous in (17) can be assumed to be person marked since person was still paradigmatic in the Hellenistic reflexive, with overt person distinctions found in the singular, for example.

20. For example, the reflexive can be modified by an adjective in the manner that nouns are modified, with the adjective between the definite article and the nominal, but not in the manner that pronouns are modified, with the definite article and adjective after the nominal; for complete details and further relevant arguments, see Joseph & Philippaki-Warburton (1986: sections 1.6 and 2.1.2.2).

21. Optionally, the reflexive form itself can occur as a plural if the antecedent is a plural noun phrase, as in :

i. emis kitazume tus eaftus mas
we/NOM watch/1PL the-selves/ACC.PL our
'We are looking at ourselves'.

In such a case, the reflexive form shows agreement in number with the antecedent, but this agreement does not encode person on the reflexive form nor is it obligatory.

22. For example, in Modern Greek written documents (books, magazines, newspapers, etc.)--a much larger "corpus" than is available for Medieval Greek, especially Medieival Greek of the 10th to 14th centuries, i.e. near when the Reflexive change occurred--I have found only two examples of Raising cum Reflexive sentences. Native speaker judgments of the form that such sentences must take, however, are remarkably uniform, all agreeing that if raising occurs and if the conditions for the appearance of the reflexive form are met, the complement clause must have third person agreement on the verb. These facts have been discussed briefly in Joseph (1978: Chapter 10; 1979), and in some detail in Joseph & Perlmutter (1979).

23. There is some controversy on this point. Ingria 1981 has drawn attention to some sentences which appear to involve raising of nonsubjects. My informants in general were most unhappy with such sentences, hence my conclusion that Raising is restricted to subjects. Moreover, the only sentences with apparent nonsubject raising that my informants accepted were those involving perception verbs, a class which has proven notoriously resistant to analysis in many languages. Ingria himself, furthermore, argues that the structures in question are not Raising structures, a conclusion I support for the nonsubject instances but not for those involving subjects. See footnote 17 for references on some of the controversy surrounding this construction.

24. See, for example, the discussion in Perlmutter & Soames 1979.

25. I am adopting here an analysis of these rules in which they do not apply in "successive cyclic" fashion.

26. One possibility is that rules that change grammatical relations are necessarily cycle (a principle to this effect was propsoed by Perlmutter and Postal in 1974 lectures). However, there are rules in English that seem to have the effect of changing grammatical relations that do not however "feed" clearly cyclic rules such as (English) Passive and thus seem not to be in the cycle. An example is the Benefactive --> Direct Object rule which produces sentences such as I baked Mary a cake; for many speakers, Mary cannot be passivized (i.e. *Mary was baked a cake (by me)), a fact which would be accounted for if Benefactive --> Direct Object were a noncyclic rule. Such an analysis would mean giving up a typology of rule cycle-type based on effect on grammatical relations.

27. I realize that actually proving that Ancient Greek Reflexivization is cyclic is an impossible task, given the unavailability of all of the relevant data that could be brought to bear on the matter; what is presented here, then, is at best a plausible account that is consistent with known facts.

28. These sentences are much more acceptable if neither Raising nor Reflexivization occur, inasmuch as Raising is somewhat limited in Modern Greek and other means of expressing reflexivity (e.g. through the use of mediopassive verbal morphology) are not only available but generally preferred. Nonetheless, there is no doubt that these rules are part of the grammar of Modern Greek, so that their interaction and application in the same "derivation" is possible in principle. Moreover, although somewhat unusual in nature (note the awkwardness of the English translation), all Greek speakers that I have consulted with on this and related sentences--some 20 in all from various parts of Greece and from a variety of backgrounds--have uniform judgments.

29. This result and related conclusions on the empirical content of the cycle as a syntactic construct have been discussed in Joseph and Perlmutter 1979; I hope to make the results more readily available in a forthcoming monographic study of Raising in Greek.

30. I have carried out about as thorough a search as possible through the vernacular texts of Medieval Greek, covering literally thousands of pages of documents; see Joseph (1978: Chapter 1) for a discussion of the methodology and references regarding the texts examined.

References

Chomsky, Noam. (1973) "Conditions on Transformations", in S. Anderson & P. Kiparsky, A Festschrift for Morris Halle. New York: Holt, Reinhart and Winston, pp. 323-386.

Coopmans, P. (1984) "Surface Word-Order Typology and Universal Grammar", Language 60.55-69.

Culicover, P. (1984) Review Article on Baker & McCarthy (eds.) The Logical Problem of Language Acquisition, Language 60.115-122.

Haiman, J. (1974) Targets and Syntactic Change. The Hague: Mouton Co.

Hawkins, J. (1979) "Implicational Universals as Predictors of Word Order Change", Language 55.616-648.

_____. (1980) "On Implicational and Distributional Universals of Word Order", Journal of Linguitics 16.193-235.

_____. (1982a) "Cross-Category Harmony, X-Bar and the Predictions of Markedness", Jounal of Linguistics 18.1-35.

_____. (1982b) "Language Universals and the Logic of Historical Reconstruction", Linguistics 20.367-390.

_____. (1983) Word Order Universals. New York: Academic Press.

Ingria, R. (1981) Sentential Complementation in Modern Greek, Ph. D. Dissertation, MIT.

Joseph, B. (1976) "Raising in Modern Greek: A Copying Process?", in J. Hankamer and J. Aissen (eds.), 290r*: Harvard Studies in Syntax and Semantics 2.242-278.

_____. (1978) Morphology and Universals in Syntactic Change: Evidence from Medieval and Modern Greek, Ph. D. Dissertation, Harvard University (printed and distributed by Indiana University Linguistics Club, Bloomington (IN)).

_____. (1979) "Raising to Oblique in Greek", Proceedings of the Fifth Meeting of the Berkeley Linguistic Society (BLS 5), Department of Linguistics, University of California, Berkeley, pp. 114-128.

_____. (1980) "Linguistic Universals and Syntactic Change", Language 56.845-870.

_____. (1983) The Synchrony and Diachrony of the Balkan Infinitive. A Study in Areal, General, and Histroical Linguistics, Cambridge University Press, Cambridge (Cambridge Studies in Linguistics, Supplementary Series).

_____. & D. Perlmutter. (1979) "The Empirical Content of the Cyclical Theory of Grammar", Paper presented at Annual Meeting of Canadian Linguistic Association, Saskatoon, May 1979.

_____. & I. Philippaki-Warburton. (1986) Modern Greek. London: Croom Helm Publishers.

Kakouriotis, A. (1979) "Raising in Modern Greek", Lingua 52.157-177.

Lieber, R. (1979) "The English Passive: An Argument for Historical Rule Stability", Linguistic Inquiry 10.667-688.

Lightfoot, D. (1974) "The Diachronic Analysis of Englsih Modals", in J. Anderson & C. Jones, eds., Historical Linguistics (Proceedings of the First International Conference on Historical Linguistics). Amsterdam: North Holland.

_____. (1976) "The Base Component as a Locus of Syntactic Change", in W. Christie, ed., Current Progress in Historical Linguistics. Amsterdam: North Holland.

_____. (1979a) Principles of Diachronic Syntax. Cambridge: Cambridge University Press (Cambridge Studies in Linuistics 23).

_____. (1979b) "Rule Classes and Syntacti Change", Linguistic Inquiry 10.83-108.

Marlett, S. (1976) Copy-Raising in Koine Greek, Master's Thesis, University of North Dakota.

Martinet, A. (1953) "Function, Structure, and Sound Change", Word 8. 1-32.

Payne, D. (1985) "Review of Hawkins Word order Universals", Language 61.462-466.

Perlmutter, D. (1983) Studies in Relational Grammar I. University of Chicago Press.

_____. & C. Rosen. (1984) Studies in Relational Grammar II. University of Chicago Press.

_____. & S. Soames. (1979) Syntactic Argumentation and the Structure of English. University of California Press.

Philippaki-Warburton, I. (1979) "O metashimatistikos kanonas tis anipsosis sta nea elinika" ["The Transformational Rule of Raising in Modern Greek"], Epistimoniki Epetiris Filosofikis Sholis Panepistimiu Thesalonikis 18.489-526.

Postal, P. (1986) Studies of Passive Clauses. (SUNY Series in Linguistics) Albany: State University of New York Press.

Russom, J. (1982) "An Examination of the Evidence for OE Indirect Passives", Linguistic Inquiry 13.677-680.

Smith, N. (1981) "Consistency, Markedness, and Language Change: On the Notion 'Consistent Change'", Journal of Linguistics 17.39-54.

OSU WPL 34 (1986) 56-83

Loss of Nominal Case Endings in the Modern Arabic Sedentary Dialects*

Ann M. Miller

1.0. Introduction

Classical Arabic (CA), which is widely accepted as representing pre-Islamic Arabic speech and therefore the ancestor language of the present-day Arabic dialects (cf. Birkeland 1952; Blau 1961, 1965, 1966-67; Ferguson 1959; Fück 1955), had nominal case endings, while the modern sedentary (non-Bedouin --generally, urban) dialects do not have these endings. Since the modern sedentary dialects differ from each other in a number of ways--even to the extent that a number of them are not mutually intelligible--the question arises as to how all these dialects came to have in common the lack of nominal case endings. This paper examines some of the evidence that has been brought to bear on this question, proposes some different analyses, and evaluates several of the existing theories in light of the new analyses.

Unless otherwise noted, the transcription used here is phonemic and uses symbols of the International Phonetic Alphabet. A dot under the consonants t, d, s, and z (ṭ, ḍ, ṣ, ẓ) indicates pharyngealization. A dot under the consonant h (ḥ) indicates a voiceless pharyngeal fricative.

2.0. The Classical Situation (Nominal Case Endings) and the Modern Situation (Reanalyzed Remnants)

The CA nominal case endings and their modern reflexes are shown in Table 1 below. In addition, the markers for feminine gender (-at 'FEM SG' and -āt 'FEM PL') and the marker for indefiniteness (-n--called 'nunation' in English, 'tanwīn' in Arabic) are shown. CA words which are definite do not include the indefinite -n but, rather, end with the vowel which marks the case ending (-u 'NOM', -i 'GEN', or -a 'ACC') or with the dual or regular masculine plural endings. The parentheses around the t in the modern feminine singular marker -at indicate that the t is pronounced only in certain environments. These environments are those in which the feminine marker is followed by a pronoun or a noun which is in a possessive relationship to the noun--the latter called 'construct state' in English, 'iḍāfa' in Arabic. In Levantine Arabic, for example, 'university', lit. 'university-FEM SG' (/jāmʕ-at/) is pronounced [jāmʕ-ə]; 'his university', lit. 'university-FEM SG-his' is [jāmʕ-It-u]; 'her university', lit. 'university-FEM SG-her' is [jāmʕ-It-he]; 'Yarmouk University', lit. 'University-FEM SG-Yarmouk' is [jāmʕ-It yarmūk]; and 'The University of Jordan', lit. 'University-FEM SG DEF-Jordan' is [jāmʕ-It Il-ʔurdun]. Note that the t in the feminine marker for the modern dual (where pronounced) and the modern regular plural is always pronounced since it is followed by a suffix (-īn) which is closely connected to it.

Table 1: Classical Arabic Case Endings and their Modern Remnants

	CLASSICAL ARABIC				MODERN ARABIC		
	Gender	Case	Indefinite-ness		Gender	Case	Indefinite-ness
SINGULAR	(F)	(M&F)	(M&F)		(F)	(M&F)	
Nominative	-at	-u	-n				
Genitive	-at	-i	-n	>	-a(t)	ø	ø
Accusative	-at	-a	-n				
DUAL	(F)	(M&F)	(M&F)		(F)	(M&F)	(M&F)
Nominative	-at	-āni	ø		-at/ø	-ayn/ø	ø
Genitive	-at	-ayni	ø	>	(Most dialects have ø, and		
Accusative	-at	-ayni	ø		plural has replaced dual)		
REGULAR PLURAL	(F)	(M)/(F)	(F)		(F)	(M)/(F)	(F)
Nominative	-āt	-ūna/-u	-n				
Genitive	-āt	-īna/-i	-n	>	-āt	-īn/ø	ø
Accusative	-āt	-īna/-i	-n				
IRREGULAR (BROKEN) PLURAL = STEM CHANGING							
		(M&F)	(M&F)			(M&F)	(M&F)
Nominative	ø	-u	-n				
Genitive	ø	-i	-n	>	ø	ø	ø
Accusative	ø	-a	-n				

As Table 1 shows, nouns in the modern dialects have generally undergone four changes from CA in phonology and morphology:

(1, 2) Phonology: unconditioned loss of indefinite -n and -V, and conditioned loss of feminine singular -t.
(3, 4) Morphology: merger of the nominative marker with the genitive/accusative marker in the regular plural and--in those dialects that retain it--the dual.

These changes have resulted in the sedentary dialects losing their nominal case distinctions.

It is unclear, however, how all these dialects have come to have this same change in common since the dialects are spread out over a vast area. Several theories have been advanced which specifically account for this phenomenon by postulating phonological and morphological changes that led to it. Prominent among them are those of Birkeland (1952), Cantineau (1953), and Blau (1961, 1965, 1966-67), all outlined below.

2.1. Birkeland's Theory

Harris Birkeland (1952), drawing on the observation that Classical Arabic had pausal (citation) forms which were essentially like the modern forms (except for the nominative/oblique merger), took these forms as the origin of the modern dialectal forms. That is, in CA -(t)V̆(n)ø in singular and broken plural forms in context (non-pause) position became ø in pause position (in isolation and sentence finally) in the nominative and genitive, and it became -ā in the accusative. Furthermore, some Old Arabic dialects

had pausal forms which ended in ∅ for all the cases. Birkeland proposed that these reduced pausal forms of these old dialects were then generalized to context position in a later stage of the dialects so that forms representing more categories replaced forms representing fewer categories: the earlier system with one form representing each of the nominative, genitive, and accusative cases gave way to a system with one form --∅-- representing all three cases. Birkeland stated that this conclusion is the only one possible because:

(1) we know that CA and some old dialects had both context forms and pausal forms;
(2) the modern sedentary dialects have only pausal forms, with context forms as relics in places that could not have pausal forms (the construct state, or iḍāfa);
(3) therefore, the form that survived had to have replaced the lost form.

Even though this conclusion is not explicit as to how the replacement happened, it is a plausible explanation of the changes in nouns that took place between CA and the modern dialects.

2.2. Cantineau's Theory

Jean Cantineau (1953) proposed that the loss of case endings was brought about by a phonetic sound change which dropped short final vowels, plus a morphological rebuilding of the case system, in the following steps.

(1) Short vowels (especially u and i) were weakened and so were subject to loss in open syllables. Therefore, first the nominative marker -u became ∅, and then the genitive marker -i became ∅. After these changes, only the accusative marker -a remained.
(2) The case system underwent morphological rebuilding to lose the nominative and genitive distinctions in indefinite nouns, too (by analogy to definite nouns): -un became ∅, and -in became ∅.
(3) A phonetic sound change made context -a and pause -ā (<-an) become ∅. After this change, context -an was the only case ending left.
(4) Then -an in context became ∅ due to morphological rebuilding (by analogy to the other forms which had ∅ endings already).

2.3. Blau's Theory

Joshua Blau (1961, 1965, 1966-67) maintained that the modern Arabic dialects grew out of Middle Arabic dialects which diverged from CA as CA spread outside the Arabian Peninsula during the Islamic conquests (c.a. 632-800 A.D.). These new dialects differed from each other because they developed in different towns, but they all lost case (and mood) endings due to (1) the influence of the foreign languages which did not have case endings, (2) the stress changing from weakly centralizing to strongly centralizing,[1] and (3) the generalization of pausal forms to context position. He argued that these changes occurred in the following steps.

(1) Short vowels in open syllables (especially word finally) were

weakened and therefore tended to drop. u and i dropped first because they were weaker than a. This resulted in nominative and genitive definite singular nouns, feminine sound (regular) plurals, and broken (irregular) plurals losing -u and -i.

(2) Nominative and genitive pause forms were extended to context, so that -un and -in became ∅.

(3) Word final long vowels became short so that pausal accusative -ā from context -an became -a.

(4) n was weakened and then dropped in open (especially final) syllables, so accusative -a became ∅. At this stage, -an in context was the only vestige of the former case markers left, no longer signifying case since the system had broken down so much.

(5) Accusative pausal forms (with ∅ ending) were optionally extended to context, so that no final case markers were left except optionally.

(6) The oblique case markers of the dual (-ayn) and the masculine sound (regular) plural (-īn) replaced the nominative markers (-ān and -ūn, respectively), since there was no longer a need to distinguish cases.

3.0. Evidence Which Illuminates These Theories

All of these theories deal with plausible types of changes, and so, since they are not mutually exclusive, it is possible that any or--as Blau argues--all of the factors which they propose could have contributed to the loss of case endings in the Arabic dialects. The task, then, is to find evidence that sheds light on what probably occurred, so that the amount of speculation necessary about what possibly occurred can be minimized.

There is a body of documents available which provides such evidence and which scholars in general--including those mentioned above (except Blau)--had not considered when developing their theories.[2] These are the writings of non-Arabs during the first five or so centuries of Arab rule (approximately the 8th through the 12th centuries A.D.). Blau (1961, 1965, 1966-67), who has analyzed hundreds of these writings, maintains that they provide information about characteristics of colloquial Arabic immediately following the Islamic conquests. As such, they are the oldest documents available which reveal the colloquial speech after the conquests provided the opportunity for extensive changes in Arabic to take place, due to the intermingling of Arabs from different areas in military campaigns and settlements and to the learning of Arabic by the conquered non-Arabs. They thus reveal a stage of Arabic which is intermediate between Classical Arabic, which had case endings, and the later stage of dialectal Arabic which does not have case endings (Blau's 'Modern Arabic'). Blau termed this intermediate stage 'Middle Arabic (MA).' As an intermediate stage, MA provides information about some of the steps the language went through as it changed from the CA type to the modern dialectal type.

These texts are written in CA, which was the standard written language, and the characteristics of MA are revealed in them as deviations from CA. Blau points out that there are no known texts written in colloquial MA, so the best that can be done to ascertain the traits of MA is to analyze texts of CA which contain deviations. He states that these texts are very

revealing sources of MA since they contain numerous deviations. He argues that such deviations represent either intrusions from the spoken language or hyper- or hypo-corrections since the writers were generally trying to write in the prestigious standard language (CA).

The deviations appear almost exclusively in manuscripts written by Jews and Christians who wrote (usually copied) mainly religious texts in their new language--Arabic. Blau notes that while a few colloquialisms occur in official Muslim papyri of this time, they do not occur often because, as the language of their religion, CA was an extremely high ideal for Arabs. Consequently, Arabs were very careful not to let many colloquialisms enter their writing, while non-Arabs were either not as careful or not as able since CA was not such a high ideal or as familiar for them. Even so, Blau points out that the few colloquialisms which occur in Arab papyri and poetry at the beginning of the 8th century A.D. have the same basic characteristics as those which occur in non-Arab texts. Therefore it can be assumed that MA was in use as early as this and that the Arabic spoken by Arabs at this time had the same basic characteristics as that spoken by non-Arabs and revealed in their writings.[3]

According to Blau (1961, 1966-67), among the non-Arabs, the texts which reveal the most about the spoken language of this time are those written by Christians in Southern Palestine for other Christians because there are many more texts available from this area than from the other areas which produced such texts. Furthermore, these writings include the earliest dated documents which include numerous examples of MA and numerous manuscripts which were written in the monasteries there in the second half of the 9th and the 10th centuries. They also include a number of undated manuscripts with numerous examples of MA for which there is evidence that they were written there in the 8th century--some as early as the beginning of the century. Most of these are translations from Greek and Syriac, but some are originals in Arabic, showing that the native non-Arabs did, indeed, produce this type of writing. As Blau points out, the dialect characteristics revealed in these documents are not homogeneous with the characteristics revealed in documents from other areas, other religions, or other times. However, his studies have shown that the basic features of all these different dialects are the same, and so Southern Palestinian Christian Arabic--or Arabic of Southern Palestine, abbreviated ASP by Blau 1966-67--can reasonably be used to represent MA as a whole, while also noting the deviations in the documents which represent only ASP or only the particular copyist. Blau (1966-67) does just this, and so the present study looks at the ASP deviations which Blau indicates are also common to other MA dialects.

Blau (1961, 1966-67) notes that precautions must be taken when analyzing MA texts because some of the deviations from CA do not represent the spoken Arabic of the time. For example, a number of the deviations are pseudo-corrections, which are a mixture of standard and colloquial features, resulting from the writers trying to use CA but not always applying its rules correctly. Types of pseudo-corrections which are found in the texts include malapropisms (such as writing lasiyyamā for lā siyyamā 'especially'; Blau 1966-67: 50), use of CA forms where they are not appropriate (called 'hyper-' or 'over-correction'--such as use of the prestigious nominative case where the less prestigious oblique case is appropriate; Blau 1966-67:

51), and mixtures of MA forms with CA forms (called 'hypo-' or 'half-correction'--such as use of a dual verb before a dual subject, when CA used a singular verb before a dual subject, and MA used a plural verb before a dual subject; Blau 1966-67: 51). Blau notes that the ASP texts also show influences from the other language spoken in the area--Aramaic--as well as loan translations from the languages that many of the texts were originally written in--Greek and Syriac. The texts also show influences from CA spelling (such as usually spelling words which had CA ظ or ض with their respective CA letters even though these sounds had probably merged in ASP; Blau 1966-67: 56, 113-114) and from traditional literary features which had disappeared from the spoken language (such as following an imperfect verb which ended in a long vowel with the symbol for -n when the dialectal pronunciation no longer included the -n; Blau 1966-67: 57). Therefore, in order to identify the true MA features from these texts and weed out the pseudo-corrections and other deviations from CA which did not represent influences from colloquial Arabic, Blau (1965, 1966-67) listed in his studies of Judaeo-Arabic and ASP texts only those features which occurred in a number of the texts as reliable features of MA, because they recurred. The present study relies only on these recurrent MA features of ASP which Blau compiled.

4.0. What These Texts Show about the Loss of Nominal Case Endings

Blau's (1966-67) compilation of a grammar of Christian Arabic based on his analysis of numerous grammatical characteristics of the Southern Palestinian texts includes a number of conclusions about the historical changes that the language underwent to reach this stage of Middle Arabic. A reanalysis of the data he considered points to some additional conclusions, some different conclusions, and some of the same conclusions, as discussed below.

4.1. A Stress Shift Could Have Occurred

Blau notes that while long vowels are generally indicated in these texts, short vowels generally are not, making it difficult to draw conclusions about ASP based on the occurrence or nonoccurrence of vowels. However, the places where vowels are indicated show that some of the vowels (short vowels more than long vowels) were sometimes written with symbols which indicated a different vowel quality than the vowels had in CA, and that long vowels were often shortened in final open syllables and short vowels were often dropped in open unstressed (especially final) syllables. The changes in short vowels are shown mainly by an 8th century fragment of Psalm 78 which is written all in Greek letters and includes the original Greek text and a translation into Arabic. Since it is written in Greek letters, it indicates all the Arabic vowels, including the short vowels--which the Arabic script generally does not indicate. It thus provides a rare window on the full vocalization of Arabic at this time.

A reanalysis of the data cited by Blau (1966-67) supports his conclusions (p. 44) that these general trends occurred. The fact that the data bears out his conclusion that 'the quality of the short vowels was rather inconstant' supports his subsequent conclusion that the vowels in ASP 'were weakened, thus becoming liable to change and elimination.' The inconstancy of ASP's vowels is shown in the examples that Blau (1966-67: 63-65) cites of ASP letters which represent different vowel qualities from CA, listed below

in Table 2.

Table 2: ASP Words with Vowel Qualities that Differ from CA (from Blau 1966-67: 63-65)

Number of instances	ASP Vowel (underlined)	for CA Vowel (underlined)	Meaning
8	e	a	
(1-3)	ⲟⲩⲗⲉⲩⲇⲓⲩ̀	wa-l-ʔawdiya and-DEF-streams	'and the streams'
(4)	ⲁⲩⲃⲁⲗ	laʕall	'perhaps'
(5)	ⲩⲉⲙ.ⲇⲓⲣ	ya-qdir 3MASC SG IMPERF can	'can'
(6-7)	ⲉϫⲧⲩⲃ.ⲁⲗⲉⲧ	ištaʕal-at kindle/PASSIVE 3FEM SG PERF	'was kindled'
(8)	ⲫⲁⲥⲉⲗⲉⲧ	fa-sāl-at and gush out 3FEM SG PERF	'and it gushed out'
2	e~a (same text)	a	
(1)	ⲥⲉⲙⲁ ~ ⲥⲁⲙⲁ	samā	'heaven'
(2)	ⲗⲩ̀ⲙ ~ ⲗⲁⲩ̀ⲙ	lą-hum to-them	'to them'
(3	u	a)*	
(1)	muqadira	ma-qadir-a noun-can-FEM SG	'ability'
(2)	yuḍrub	ya-ḍrub 3MASC SG IMPERF-beat	'he will beat'
(3)	yuṣīr	ya-ṣīr 3MASC SG IMPERF-become	'he will become'
2	e	i	
(1)	ⲃⲩⲇⲁ	jiddan	'much'
(2)	ⲩϫⲧⲉⲃ.ⲁⲗⲉⲧ	ištaʕal-at kindle/PASSIVE 3FEM SG PERF	'was kindled'
2	u	i	
(1)	sulm	silm	'peace'
(2)	mush	mish	'haircloth'
1	i	u	
(1)	ⲧ.ⲏ.ⲟⲩⲣ	ṭuyūr	'fowels'
3	ē (written as ā in Arabic)	ā	
(1)	mamrē	mamrā	'Mamre'
(2)	ⲫⲁⲥⲉⲯⲉⲧ	fa-sāl-at and gush out 3FEM SG PERF	'and it gushed out'
(3)	ⲗⲓⲇⲉⲯⲗⲓⲕ	liḏālik	'therefore'

*The parentheses around the listing of ASP u written for CA a indicate that these instances may represent morphological, rather than phonetic,

substitution. This is so because all of the examples that Blau cites exhibit the substitution in a prefix: one instance of mu- for ma- (prefixes for verbal nouns), and two instances of yu- for ya- (prefixes for imperfect active verbs). Since these prefixes which contain the u occur in Arabic—and frequently—it would not be surprising if the non-native speakers of Arabic occasionally mixed up the prefixes which contained u and a. If ASP u for CA a were a phonological change, one would expect to also find it in environments other than those which are morphologically defined (here, prefixes). Therefore, it seems that the u for a substitutions noted by Blau and listed here should not be included in data showing that vowel quality in ASP was inconstant.

The data in Table 2 shows that CA a, i, u, and ā were subject to phonological change in ASP and that, in general, the change was centralization: a>e, i>e, and ā>ē. Also, occasionally i and u were interchanged. Centralization could have been a reason for this, too, if the pronunciation of these vowels diverged from peripheral toward central so that hearers perceived them as falling within the opposite phoneme boundary. All these changes point to a situation in which these four vowels varied from their CA pronunciations, at least sometimes, enough that ASP hearers (including writers) perceived them as different vowels, and then ASP writers wrote them as the different vowels. In such a situation, it would not be unusual that fewer of the long vowels varied in their pronunciations than the short vowels did (as this data shows—only ā; not ī or ū; but a, i, and u) since their longer duration would have made them more resistant to centralization, both in production and in perception.

Along with this inconstancy of vowel quality, the data listed by Blau also indicates that, in contrast to CA, long vowels were shortened in final open syllables, and short vowels were deleted in open unstressed syllables—especially word finally. Some of the evidence cited by Blau in support of the first claim is that words which end in CA -ā are sometimes written with -a in ASP, and CA -ī is sometimes written as -i in ASP. The second claim is supported by Blau's report that a symbol indicating the lack of a vowel (Arabic sukūn, symbolized ˚) following the consonant it is written above is sometimes written in ASP at the ends of words which ended in a short vowel in CA. The loss of short vowels in open unstressed syllables in ASP is further supported by Blau's observation that a symbol indicating glottal stop followed by a vowel (Arabic ʔālif, symbolized ˀ) is sometimes added before an initial consonant that was followed by a short vowel in an open unstressed syllable in CA. Blau reasons that a vowel was added before the initial consonant of the word because the unstressed vowel following this consonant had been dropped. The vowel was inserted, apparently, in order to break up the consonant cluster which resulted when the unstressed vowel was dropped—a phenomenon which is common in Arabic. For example, CA ʕlyhm ([ʕalayhim]) was written in ASP as ʔʕlyhm ([ʔaʕlayhim]).

Blau (1969: 221, 1965: 45) states that the changes in the vowels in ASP described above played an important role in bringing about the loss of case endings and that a factor in bringing about these vowel changes was a change in stress. He claims that CA must have had weakly centralizing stress (see Footnote 1) because short vowels were preserved in open unstressed syllables, but that the stress must have shifted to strongly centralizing in ASP because

short vowels were then blurred in open unstressed syllables, as described above.[4] While it seems likely that the changes in vowels (especially loss of final short vowels) contributed greatly to the loss of the case endings -- some of which were marked solely by particular final short vowels -- it is not clear whether a change of stress did or did not bring about these vowel changes.

The conclusion that a stress shift occurred is consistent with the facts, and so it is a possible explanation for them. It is widely accepted that vowels which get centralized (reduced) are unstressed and often occur in open syllables, especially at the ends of words. Since centralization weakens the vowels (makes them less perceptually distinct), such vowels are often subsequently lost altogether. Therefore, it would be expected that if the stress in Arabic had changed from CA to MA in such a way as to favor centralization of vowels more than it had before, then more vowels than before would show centralization and possibly total loss in these environments. Since this prediction describes the phenomena exhibited for the vowels which occur in the ASP texts, the conclusion could be reached for this stage of Arabic that a shift in the type of stress had occurred.

However, such a conclusion is not required by the facts. Vowel centralization can occur whenever a syllable is unstressed; it does not need to be preceded by a shift in stress. The syllables in which the vowels were reduced or lost in MA could also have been unstressed in CA but not have undergone vowel reduction or loss yet. If this was the case (and there is no evidence that it was not the case), then ASP would simply be the stage at which the vowel changes occurred, after the impetus for the changes was set up at an earlier stage. Therefore, since such a situation does not require positing that a shift in stress occurred between CA and MA, the vowel phenomena do not show that there had necessarily been a shift in stress; they only show that there could have been a shift in stress.

So these data show that Cantineau's and Blau's theories that the loss of case endings began with a shift in stress could be right but may not be. The fact that they give enough information to show that these theories could be correct is a step forward from the argumentation supplied by Cantineau and Blau, who extrapolated their conclusions from only a few facts. The fact that these data show that there is not enough information to confirm these theories is also a step forward, since Cantineau and Blau both assert that a stress shift did occur, implying that the evidence definitely supportssuch a conclusion.

4.2. Nominal Case Endings May Not Have Been Lost Completely Yet

The ASP texts give evidence that the nominal case endings had been lost at the ends of words by this time, supporting Blau's (1961, 1965, 1966-67) claim that the case endings had completely disappeared by the time of ASP. However, contrary to Blau's claim, the ASP texts also give evidence that the case endings may not yet have disappeared when followed by a pronoun suffix. The evidence supporting this situation of partial preservation of the case endings at this time is examined below.

4.2.1. Case Endings at the Ends of Words

Blau (1967: 317-318) reports that, except for one word (χεδδα) for ǰiddan 'much', where -a for -n- marks the accusative case--discussed below in Section 4.2.4), the Greek/Arabic fragment of Psalm 78 exhibits no case endings at the ends of words. The examples that Blau gives are listed below. Here, and in the rest of this paper unless otherwise noted, underlining of a blank space in the ASP text indicates the place where a CA letter would have occurred. Underlining in the corresponding CA word shows the CA letters that are not indicated in the ASP text.

	ASP:	for CA:
(1)	ωαχουφ3-....	wa-xubz-a-n ...mā?id-at-a-n
	μαίδεύ---	and-bread-ACC-INDEF...table-FEM SG-ACC-INDEF
		'and bread...table'
(2)	λυχουμ--	luḥūm-a-n
		meat-ACC-INDEF
		'meat'

Case endings are not listed in the Greek/Arabic psalm even when the noun occurs in the 'construct state' ('iḍāfa construction' in Arabic)--a syntactic construction made up of a series of nouns which indicate possession of N_1 by N_2 and--if three nouns occur--N_2 by N_3. In this construction, the nouns are very closely tied together and therefore--except the final word--would not be pronounced in their pause forms in CA. Blau states that in ASP, however, such nouns are written in their pause forms, and he gives the following example.

ASP:	for CA:
ωαμίολ..ραμλ-.	wa-miθl-i raml-i-l-buḥūr
ελβουχουρ	and-as-GEN sand-GEN-DEF-sea
	'and as the sand of the sea'

In this example, the case vowel of CA miθli was not written in the ASP text, and Blau says that the case vowel of CA ramli also was not written in the ASP. Since Greek ε- which occurs in this text after the Greek for raml --is a vowel, though, this could be the i of ramli. This possibility is not likely, since the dots in the Greek rendering of ASP apparently indicate word boundaries, but it should be considered and investigated further. In any case, this example shows that ASP dropped at least some case endings in this construction. Since this position is so resistant to deletion, this is strong evidence that ASP had either, as Blau claims, totally lost the case distinctions which formerly occurred at the ends of words (if the Greek ε was not the i from ramli), or nearly lost these distinctions (if the ε was the i from ramli).

Two other examples which Blau (1967: 320) cites as evidence that the cases had been lost in ASP actually show that the form which marked the cases in CA did not always disappear--sometimes it just ceased to carry out

its former function. In these two examples, listed below, the noun ʔax 'brother', which is in the construct state, ends in the form of a CA case ending which is an incorrect ending for this context: ū in ʔax-ū ('brother-NOM') signalled the nominative case in CA, but this construction required a genitive marker (-ī) on this noun.

ASP:

(1) l(i) y(a)ʕqūb ʔ(a)x ū r(u)b nā
to James brother NOM lord our

'to James, the brother of our lord'

(2) r(i)sāl-a y(a)ʕqūb ʔ(a)x ū r(u)b nā
epistle-FEM.SG James brother-NOM lord our

'the epistle of James, the brother of our lord'

Therefore, these endings, while retained in form at the end of the word, appear to no longer be functional as case markers. This situation is discussed in detail in the next section.

4.2.2. Case Endings Before Pronoun Suffixes

Blau (1967: 318 n.3) points out that sometimes the CA case endings were omitted in ASP before attached pronoun suffixes, and he claims that these instances represent the actual ASP usage. The two examples he cites of this type of omission, from the Greek/Arabic Psalm 78, are:

ASP:	for CA:
(1) χε.ύ.οετ..ύμ	šahw-at-a-hum desire-FEM.SG.ACC their (MASC) 'their desire'
(2) καδ.ο..ύυ	quds-i hī sanctuary-GEN his 'of his sanctuary' (no overt preposition)

He notes that at other times the case endings occurred in this position in ASP, and he claims that these instances were not the general usage but were due to the influence of CA - that is, that they were hyper corrections. The two examples he gives of this are the following, with the case endings underlined.

ASP:	for CA:
(1) βι.λαυ.οάγ.ι̣ύμ	bi ʔawθān i-him with-idols-GEN their (MASC) 'with their idols'

(2) بِ.مِنْ.خُوتِ. bi-mnxūt-āt-i-him
with-graven image-FEM PL-GEN-their (MASC)

'with their graven images'

If these occurrences of case endings are hyper-corrections, it is curious that they occur only before a pronoun suffix and never at the end of a word. If the writer was correcting his Arabic according to the rules of CA, it would be expected that he would have at least occasionally written case endings in the most obvious place they occur in CA--at the ends of words. Since he did not do this, it raises the question of whether the instances of case endings before the pronoun suffixes are, indeed, instances of hyper-correction.

The alternative is that these case endings represent the actual usage of the time. Perhaps case ending vowels had not been totally lost yet but were still pronounced--at least sometimes--when they were not at the end of a word. This is plausible, since such vowels would have been protected by the suffixes which attached after them, so they would have been less susceptible to changes that affected the ends of words than vowels which came at the ends of words would have been.

Additional evidence in support of this hypothesis comes from the examples Blau (1967: 318-321) gives of vowels that represent cases in CA which occur in ASP texts after ʔab 'father' and ʔax 'brother.' Blau cites 124 instances where this happens in a number of manuscripts. Of the 12 examples that he writes out fully (which include 26 instances of ʔab or ʔax), all but two (those listed above) have pronoun suffixes attached after the vowels. This is a substantial increase over the examples noted by Blau in both the number of examples and the number of manuscripts in which these vowels were written in ASP. Therefore, these examples make it look more plausible than Blau indicates that at this time what had been case vowels in CA continued to be pronounced when they occurred before pronoun suffixes (as well as sometimes without the suffixes).

This evidence is not unquestionably supportive of the theory just advanced, and the theory is not without qualifications. But the possibility that CA case vowels were pronounced in ASP at least sometimes (possibly mainly before pronoun suffixes) is one of several scenarios that would explain all this evidence without leaving problematic exceptions that need to be explained as influence from CA in only limited environments--as Blau's theory does. All these explanatory theories deserve to be considered, and so the qualifications of the above theory as well as the rest of the theories are discussed below.

One of the problems that must be accounted for is that, as noted above, the case vowels which occur in the ASP texts described here are often the wrong vowels for the cases that should occur in these positions. Of the 124 examples involving ʔab and ʔax which Blau cites, almost all are examples in which the wrong case vowel (often ū) was used. The seven which are examples of the correct vowel being used are from manuscripts that Blau says are grammatically corrected (even though they also exhibit deviations from CA in the case vowels), and so he discounts them. Only the two examples first

mentioned in this section—from the Greek/Arabic psalm and one instance of ʔab in an example that includes three instances of ʔab—have the correct CA vowel for the case ending without the possibility of it having been corrected after the original writing. The theory that the case vowels were sometimes pronounced in ASP must therefore be qualified to account for many of these vowels being wrong. Three possibilities exist to account for these vowels in the theory just proposed.

One possibility is that a vowel was often pronounced in the case-marking position but that the particular vowel always or often varied randomly so that sometimes it matched the CA vowel used to mark the particular case, and sometimes it did not. Such a situation would have occurred if ASP had a rule to insert a vowel—but, for many people at least, not any particular vowel but often ŭ—after a noun in particular environments (mainly before a pronoun suffix). If this was a rule that not everyone used or that was violated occasionally, the few exceptions in the examples examined here in which a vowel was not inserted between a noun and a pronoun suffix would be explained. If this rule was sometimes extended to nouns in the construct state, the two examples examined here of a wrong case vowel being used in the construct state would be explained. In such a situation, the system of case marking would have either broken down entirely already or would have been in the process of breaking down, depending on whether some speakers still had a sense of case marking—even possibly including some rules to place the correct CA vowel in the correct position at particular times.

The second possibility is that the case vowels were pronounced only in the environments exhibited here. That is, the words ʔab 'father', ʔax 'brother', ʔawθān 'idols', and manxūtāt 'graven images'—or religious words in general—might have been preserved longer in their older forms (followed by case vowels) than other nouns in ASP were, because of having a special status as religious vocabulary and because of being repeated often in their old CA forms in religious contexts, or as the result of loss by diffusion. The forms followed by pronoun suffixes could have been preserved longer than the forms without these suffixes, due to protection by the suffixes. In this situation, ASP or some speakers of ASP could have had rules such as those described for the first possibility above, except that the environments would have been specified either for these particular words or, more generally, for religious words. This would be a situation in which the system of case marking had broken down even more than it would have for the first scenario described above, but—contrary to Blau's theory—some sense of it would still have existed.

A third scenario is also possible—that the case vowels represented by the Arabic ASP writings are indicative of only the writing system and not of the spoken language of ASP. In this situation, the vowels concerned are the long vowels attached after ʔab and ʔax, since these are the only case vowels that occur in these examples in writings done in the Arabic script. In this situation, if the vowels following ʔab and ʔax are only part of the Arabic writing system for ASP, then they likely represent an earlier pronunciation, and so they still have something to reveal about the history of spoken Arabic case endings. For the same reasons as given above, whenever the spellings of the case vowels began to not follow the CA rules, they were probably following either current pronunciations or recent pronunciations exhibited by

a sense that a vowel should be attached after a noun in particular environments). So if the case vowels had ceased to be pronounced entirely by the time of ASP, the ASP spellings show that these vowels had formerly been pronounced for a longer time before pronoun suffixes--either in religious words or generally--than in other environments. In this situation, then, the ASP spellings just continue an earlier writing tradition that placed random long vowels (often ū) in the environments in question.

There is other evidence, though, that there was some awareness of case endings at the time of ASP. This is provided by two examples that Blau (1966-67: 318 n.3) mentions from the Greek/Arabic psalm. In these, no case endings are present, but the vowel of the pronoun suffix has been changed to agree with what the vowel of the genitive case ending would have been if it had been there, as was done in CA when the genitive ending was present. That is, hū 'he' has become hī in vowel harmony with the preceding (here, missing) genitive marker -i-, as shown by the underlinings in the reproductions of Blau's examples below.

ASP:	for CA:
(1) λιζχ]χειγ.ρ_ῦ́ῑ́	li-šaʕb i-hī
	for people-GEN his
	'for his people'
(2) γαλα χαλαθ_ῑ́	ʕalā xalāṣ i hī
	in salvation-GEN his
	'in his salvation'

Blau terms this phenomenon 'remarkable' and attributes it to hypo-correction -- a mixture of ASP and CA.

But this does not have to be seen as a remarkable and unexplainable phenomenon except by appeal to the influence of CA. These words could, in fact, show the real ASP usage--that the pronoun vowel was pronounced in these words so as to represent the genitive case ending in some way. There are three possibilities for the way this could have been done.

First, the -i of the pronoun could have represented the genitive case ending directly. Such a situation could have been brought about by speakers being aware that these expressions should have a case ending but reanalyzing the ending and thinking that the case should be marked at the end of the expression rather than at the end of the noun. In such a situation, speakers would have placed the case ending at the end of the expression, replacing the original vowel of the pronoun -hū with the genitive case marker -i. This use of the genitive marker could have been fostered by speakers hearing these expressions pronounced with -i at the very end only--instead of following the noun itself. Such a pronunciation could easily have occurred if speakers elided the unstressed case vowel after the noun, e.g.: li-šaʕbihī → li-šaʕbhī. Then hearers could have reanalyzed the -ī at the end of the expression as a case marker.

Second, even if the case endings had already been lost after nouns as Blau claims the phenomenon here could be explained if speakers were still aware of cases and knew that in CA the vowel in the pronoun suffix hũ was changed to ĩ when the construction was in the genitive case. When they knew the case of an expression was genitive, then even though the expression did not have a case ending to mark it they would have changed the pronoun ending hũ to hĩ. It would have been obvious to Arabic speakers from the occurrence of the preposition in these examples that the genitive case was appropriate here, and they would have marked this case by changing hũ to hĩ.

This possibility suggests that case marking phenomena may have been preserved longer when elements in the environment made it obvious what the case was. None of the four examples cited by Blau (1966 67) in which the case endings were lost totally (without even adjustment for them in the pronoun ending) has an overt element (e.g. preposition) to signal what the case should be. On the other hand, the four examples Blau cites as exceptions to his thesis that case endings had disappeared totally in ASP (in which either the case endings were preserved or the vowel of the pronoun suffix was changed to agree with the absent case ending) were preceded by an overt element (preposition) which would signal the appropriate case. This is a small set of data, though, and so is only suggestive of a possibility rather than indicative of a probability.

Consideration of the examples Blau gives for ʔab and ʔax extends the data somewhat and provides support for this hypothesis, which was suggested above by noting that the eight examples Blau gave for case endings having disappeared altogether in ASP do not unequivocally support his claim. The data for ʔab and ʔax do not neatly fit the pattern noted for the previous eight examples, but some do, and the rest do not contradict this hypothesis. Of the 124 examples Blau gives in which a vowel different from that called for by the CA case system follows the noun, six nouns directly follow a preposition, as shown below. Since these examples are in Arabic script, short vowels are not indicated, but case endings are indicated by long vowels since these nouns have attached pronoun suffixes. The prepositions and the case vowels which are attached to the following nouns are underlined below.

(1) bnũwat ʔab ũ nã ...dʕwat nã ʔiyã h
sonship father NOM our...call our particle him

ʔab ũ nã ...w l ʔab ĩ h
father NOM our...and to father GEN his (the last case vowel is correct)

'to be sons of our father...to call him our father...and to his father GEN'

(2) mʕ ʔab ã hmã
with father ACC their (DU) (incorrect case vowel)

'with their father ACC'

(3) mʕ ʔab-ī-hmā
with father-GEN their (DU) (correct case vowel; from a manuscript which exhibits corrections)

'with their father-GEN'

(4) b-ʔax ū-hā
of brother-NOM her (incorrect case vowel)

'of her brother NOM'

(5) qdām ʔax ū h
before brother NOM his (incorrect case vowel)

'before his brother-NOM'

(6) y-tklm ʕlā ʔax-ū-h ʔaw
3MASC.SG.IMPERF-speak against brother-NOM-his and

y-dīn ʔax ū-h
3MASC.SG.IMPERF-judge brother-NOM-his (incorrect case vowel)

'he speaks against his brother-NOM and judges his brother-NOM'

Of these six examples which begin with a preposition, two follow the noun with ī (which is the correct ending in CA when the noun also has an attached pronoun suffix), one follows with an accusative marker (ā), and the other three with nominative markers (ū). The two which have the genitive marker following a preposition are further examples of the correct case ending occurring when a preposition overtly indicates the genitive case, and so they also support the hypothesis that cases remained longer in this type of situation. The four other examples mentioned here which have the wrong case ending even though they are preceded by a preposition do not provide support for this hypothesis, but they are consistent with it. These four could well show that even in this situation the sense of case marking was also breaking down or was breaking down for some people.

The fact that all the other examples which Blau gives of the wrong case vowel being used with ʔab and ʔax are instances which do not occur with a preposition to overtly signal the correct ending makes it all the more curious that the only places that the correct ending occurs in all these instances are those in which the noun directly follows a preposition. This is further support for the hypothesis suggested here. Further analysis of the data from these manuscripts needs to be done to check this hypothesis more thoroughly, but these examples at least indicate that this situation is a possibility. It should be noted that if this is borne out, it would contradict Blau's (1961: 81-82; 1966-67: 46 n.49) suggestion that the existence of prepositions was a cause for the loss of cases because they made case endings less necessary by marking one case themselves. It seems, however, that if the more detailed analysis suggested here is borne out, then a rejection of the more general conclusion reached by Blau would be warranted.

There is a third possible explanation besides Blau's for the occurrence

of the case vowel before the pronoun suffixes and for the changed vowel in the pronoun suffixes even when the genitive case ending was not present. These occurrences might have been borrowings from CA which originally contained the genitive ending _i as well as a changed pronoun vowel, but subsequently lost the genitive _i while retaining the changed pronoun vowel because it was fixed that way in these expressions. This looks reasonable for the four expressions that do not contain ʔab and ʔax (which Blau treats in a section by themselves). These four are repeated below, with the case marker and changed pronoun vowel underlined.

ASP:

(1) bi-ʔawθāu i him
with-idols GEN their (MASC)

'with their idols GEN'

(2) bi manxŭt-āt i-him
with graven image FEM PL GEN their (MASC)

'with their graven images'

(3) li ʕanʕb hī
for people ø his

'for his people'

(4) ʕalā xalāṣ hī
in salvation ø his

'in his salvation'

All of these could easily be expressions that were used repeatedly in religious ceremonies and so became fixed in a form that was closer to the original CA than everyday ASP was.

When the ʔab and ʔax data is considered in relation to this possibility, though, it does not fit in as well as the above four examples do. All of these new examples are religious in nature and could easily have been repeated often in religious ceremonies. Yet only two of them have the correct case vowel. So in these examples the original CA system was not retained as it was in the four examples above, and it appears that these examples with ʔab and ʔax would not have been fixed in their CA form as the previous four could have been. Even if these four examples were fixed in their CA form, it is still curious that the noun in each is immediately preceded by a preposition. This, again, points to the reasonableness of the previous hypothesis.

Whatever the reason, though, for the changed pronoun suffix in the expressions which do not have a preceding genitive case vowel, these expressions, along with the expressions which lost the case vowel but did not change the pronoun vowel, provide evidence about a cause of the loss of the case endings. They show that the loss of the endings cannot have been due

just to the loss of final short vowels without something causing internal changes as well—as Cantineau maintained—because these words in which the case vowel dropped did not have the vowel at the end of the word. Rather, the vowel was inside the word before the pronoun suffix, where it would have been protected from loss due to dropping of word final vowels. Instead, the loss here must have been due to either elision of unstressed vowels—discussed in Section 4.1 above—or to a generalization from other forms. If this loss was due to generalization from other forms, this suggests two possibilities.

First, the generalization could have been from pausal forms to context forms—as Birkeland and Blau maintained. In such a situation, speakers would have realized that nouns were spoken without their case vowels when they were alone or at the ends of utterances, and they could have then started pronouncing nouns inside utterances in the same way. This would no doubt have been a gradual process, and so one of the last contexts for the ø ending on nouns to be generalized to could well have been that just suggested—nouns with attached pronoun suffixes, particularly nouns used in religious ceremonies, and particularly nouns directly preceded by prepositions—which overtly indicated the appropriate case.

Second, the generalization could have come from nouns that had lost their case endings due to another reason, such as phonetic change or generalization from pause forms. In this situation, speakers would have realized that some nouns which were in context did not have case endings, so the motivation to use case endings there would have disappeared, and speakers could gradually have quit using case forms in context. Again, such a process would have been gradual, so that nouns in environments that obviously indicated their case, as described above, could well have been the last to lose their case endings.

Furthermore, the data containing ʔab and ʔax which Blau cites also provides evidence about a possible cause and a direction of the loss of case endings. In this data, as Blau notes, by far the prevalent wrong case ending is -ū, which is the nominative marker. This suggests that the nominative form was being generalized as the form for ʔab and ʔax in all positions. The reason for this is unclear, but a reasonable possibility is that there were common religious phrases that included ʔab-ū 'father-NOM', which made the nominative form of ʔab very frequent and the prevalent form in speakers' minds. If this happened, then it would have been easy for speakers to generalize ʔabū to other positions where ʔab was used. Then speakers could easily have extended the common use of the nominative case for ʔab to the closely related ʔax 'brother', using ʔaxū in most positions as well. If this happened, speakers could well have been confused about what the appropriate use of the case endings was. Such confusion could have been extended to the use of case endings throughout the whole nominal system, contributing to their loss.

It is clear that more questions are being raised than are being solved by this examination of data showing where ASP used, used incorrectly, and did not use CA case endings. A number of new possibilities have been suggested, though, by this examination, showing that more information can be gleaned from the data available than has been recognized so far by

researchers. Several of the proposals here also show that the data may provide a more coherent whole than has so far been demonstrated. For example, the possibility that case endings may have been retained longer when followed by a protective environment such as a pronoun suffix, or when preceded by an overt marker of case such as a preposition provides an explanation for facts that were previously viewed as exceptions to the apparent generalizations. In order to settle the questions raised here, a wider examination must be done of the data available with the goal of verifying or rejecting these proposals.

4.3. The Accusative Case in Singular and Broken Plural Nouns

4.3.1. The Evidence

Blau (1966-67: 323-345) describes a number of ASP usages of the indefinite accusative marker -an which occur both in accordance with and contrary to CA usage. First, in the ASP texts, accusative -an sometimes appears where it would in CA to mark the triptotic singular and broken plural, and sometimes it does not. While Blau does not discuss the frequency of occurrence of the accusative marker (except for adverbs discussed below in Section 4.3.4), he says the occurrences and nonoccurrences alternate 'freely', which indicates that there is no apparent reason (except for adverbs) for their occurrence or nonoccurrence. His conclusion is that this is evidence that the cases have already disappeared and that ASP is a mixture of MA and CA.

Second, sometimes accusative -an occurs in ASP texts where it would not in CA. One of these instances is more common than the others -to mark adverbs regardless of case (discussed below in Section 4.3.4). Of the other instances which are less widespread in the ASP texts, Blau notes that some occur in the same categories as in modern Bedouin dialects, and others do not. While stating that all of these usages probably occurred in ASP speech, Blau attributes their occurrence in the ASP texts to hyper correction. He reasons that since the most common Bedouin usage of -an which is contrary to CA usage (marking indefinite nouns followed by an attribute) is not found in the ASP texts, then there must have been no hyper-corrective factors to bring it into the texts and that, therefore, the other occurrences of -an contrary to CA usage are due to hyper-correction. He notes that most of these instances would have required the nominative case and a few of them the genitive case in CA, but he draws no conclusions from these observations.

Third, in a manuscript from the 10th century A.D., accusative -an is often used in every syntactic environment, replacing even the nominative and genitive endings that would have been appropriate in CA. Sometimes this manuscript also omits -an even where it was appropriate in CA. Blau states that this cannot be an idiosyncrasy of the manuscript or of the copyist since two later (13th century) manuscripts which are unrelated to the first also exhibit these characteristics. In his earlier work on Judaeo Arabic, Blau (1965: 210-211) wonders whether these manuscripts reveal a situation in which 'tanwīn -un and -in had already disappeared, but tanwīn -an could be used optionally in every syntactic environment....' In his later work on ASP, however, Blau (1966-67: 340 n.84) terms his earlier assumption 'rather daring' because it would force the postulation of 'a very intricate history of development' of the ASP manuscript and its two related predecessors. So

Blau makes no conclusions about this situation, either.

Once again, these facts point to more information about spoken ASP than Blau deduces. As with the data discussed above in Section 3, the data here do not show conclusively--contrary to Blau--that case endings had already been lost in ASP. The data here do show that the writers of ASP manuscripts were deviating substantially from CA rules and therefore that the case system as it existed in CA was not in spoken use any more (if it was, the writers would not have deviated from it as much as they did). But this does not mean that case endings were completely absent in speech. As with the data in the previous section, this data is consistent with the possibility that some sort of case system or some sense of a case system existed at the time of ASP, and so that possibility merits consideration.

The fact that in these writings the accusative case was used a number of times in place of the CA nominative and genitive cases suggests that the accusative case had some psychological reality for ASP speakers. That is, -an as an indefinite case marker might have still been in use enough that speakers were aware of it as a case marker and so optionally extended it to positions where they knew any indefinite case markers were used. This could have happened if the other case markers had been lost faster than the accusative marker, so that the others lost psychological reality as a whole before the accusative marker did. (The others may still have retained psychological reality in particular contexts where they were overtly marked, as suggested above for the genitive marker when preceded by a preposition.) A reason for the accusative marker being retained longer than the other case markers could have been its greater sonority and therefore perceptibility, as a low or mid vowel, than the other vowels, which are high. This explanation for the ASP data therefore supports the theories of Blau and Cantineau that -a was retained longer than -u and -i due to phonetic factors. It also supports Cantineau's and Blau's claim that at one point in the history of spoken Arabic -an was the only case marker left, while the nominative and genitive markers had already been lost so that many nouns had ø endings at this time.

Furthermore, this theory--that the accusative marker was retained longer than the other case markers and was even optionally extended to the positions of the other markers--is appealing because it can explain some problems raised by Blau and can tie together some conclusions reached from the ASP texts. First, this possibility could explain the lack in the ASP texts of -an marking a noun followed by an indefinite attribute in positions that would have called for the nominative or genitive case in CA, without having to call it a 'remarkable phenomenon' as Blau (1966-67: 329) does. If -an were being or had been generalized in speech to positions formerly occupied by only nominative or genitive markers while a sense of the case system still remained, use of -an to indicate that a noun in any position was followed by an indefinite attribute would stand out as being contrary to the CA case system. It could very well have been avoided by the ASP writers precisely because they used it in their speech but recognized it as a deviation from the preferred usage. One does not, then, have to conclude, as Blau does, that there were no hyper-corrective factors at work on this construction while there were on the other constructions in which -an appears contrary to CA usage in these texts. While possible, Blau's conclusion seems unlikely, since if hyper-correction was at work in most of these situations,

it is odd that it would never be found in one of them. The other usages of _an could easily have slipped into the writings from speech because they are less easily identifiable as contrary to CA usage since there are similar CA constructions which take the accusative case. In such a scenario, then, all the data are accounted for by the same phenomenon, rather than positing the existence of one phenomenon in most of the instances but a lack of its existence in one situation.

Second, this theory could explain the occurrences of _an in all syntactic positions in the one CA manuscript without having to posit a complicated history of this manuscript and its related manuscripts, which were written in the century before it was written. If the accusative case marker gradually spread to positions where formerly only the nominative and genitive markers were used, then this manuscript could reflect the situation in which the accusative marker had finally spread to all these positions. The two related manuscripts which were written earlier could reflect an earlier situation in which the accusative marker had not yet spread to all the other positions. Blau (1965: 211) assumes that the use of the accusative case optionally in all positions (even those in which it occurred obligatorily in CA) would have been the more archaic stage of these two, calling it 'the oldest stage of the retention of tanwīn, after the breakdown of the case system of Classical Arabic....' But if the occurrence of _an optionally in all syntactic positions is seen as the end of a process of the accusative case being generalized to other positions rather than as the beginning of the loss of tanwīn, then the use of the accusative case optionally in all positions would follow its use in some positions formerly occupied by case markers. In such a situation, the problematic ASP manuscript is no longer a problem because its structure logically comes later than the structure of its chronological predecessors, so it can credibly be seen as representing that which it intuitively seems to represent - a stage in the spoken language of ASP.

Finally, as discussed in the next section, this theory of the history of the indefinite accusative marker in singular nouns and broken plurals would provide a unified account of the history of the accusative case throughout the nominal system.

4.3.2. The Oblique Case throughout the Nominal System

Blau (1965 67: 218 226) indicates that in ASP the oblique (accusative and genitive) case had replaced the nominative case of CA in dual nouns and in masculine sound (regular) plurals. This is shown by the very frequent use of _ayn in the ASP texts where _ān was used in the nominative of CA duals, as shown by the following example.

ASP:	CA:
hāðayn l-nabyayn	hāðān l-nabyān
this/OBL DU DEF-prophet-OBL DU	this/NOM DU DEF prophet-NOM DU
'these-OBL two prophets-OBL'	'these-NOM two prophets-NOM'

The replacement of the nominative case by the oblique case is also shown by the very frequent use of _ī(n) in ASP texts where _ū(na) was used in the nominative of CA masculine sound plurals, as shown by the example below.

ASP:	CA:
y-urā l barān-īn	y-urā l barān-ūn
3MASC.SG.IMPERF-think-DEF-stranger-OBL MASC PL	3MASC.SG.IMPERF-think-DEF-stranger-NOM MASC PL
'the strangers-OBL think'	'the strangers-NOM think'

Since, according to Blau, these usages are so frequent, the conclusion that they reflect spoken ASP seems warranted. The theory proposed here that the indefinite accusative marker was generalized throughout the nominal system for singular nouns and broken plurals could be combined with Blau's observations --that the oblique marker replaced the nominative marker in dual nouns and masculine sound plurals--to yield the general hypothesis that in ASP the nominative case in nouns was replaced by an oblique case. Such a theory is appealing because it unifies what have formerly been treated as unrelated phenomena, suggesting that ASP speakers treated the whole nominal case system the same way, rather than treating its different components separately (excluding, of course, instances of analogy which were confined to specific lexical items or contexts, such as that described above in Section 4.2.2 for ?ab and ?ax).

4.3.3. Generalization of the Internal Oblique Markers

The timing of these changes is not clear from these texts, though. As described above in Section 2, researchers who have included phonetic factors in their chronology of events in the loss of case endings (cf. Blau and Cantineau) have considered such factors to have played a motivating role at the beginning of the chronology of events in the loss of the case system. Because of its neatness, it is certainly appealing to assume, as Blau does, that (1) phonetic factors and generalization of pausal forms to context brought about the loss of final short vowels, thereby doing away with the case distinctions that were marked by final short vowels only; and (2) other case distinctions were subsequently lost by analogy to the forms which had lost final case endings, since the reason for the case distinctions had become blurred with the loss of final short vowels.

However, the generalization of the oblique nominal case markers to nominative contexts in the dual and the masculine sound plural suggests another possible sequence of changes. Since the oblique markers represented two cases in CA while the nominative markers represented only one, speakers could easily have generalized the marker which represented the greatest number of cases to the positions of the marker which represented fewer cases--without needing prior dropping of final case vowels elsewhere in the system to blur the case system and trigger these changes. Such a possibility for the beginning of the loss of case endings is supported by the fact that Blau (1965: 127 n.1) mentions that the oblique case occurs twice in the Qur'ān for the nominative case in the masculine sound plural, while he states that CA preserved short vowels (Blau 1961: 213, 1965: 69, 1966-67: 43), which means that the phonetic changes did not happen until Middle Arabic. This shows that the generalization of the oblique marker could occur without being aided by the phonetic change.

If the possibility suggested here had been the motivation for the loss

of the case distinction formerly made by internal long vowels, then the chronology of events in the loss of nominal case endings would place this event as beginning first, followed shortly by the phonetic changes and generalization of pausal forms to context. In such a scenario, most of the separate events would have taken place concurrently, and loss of the nominal case distinctions in the dual and the masculine sound plural could have contributed by analogy to the loss of case markers at the ends of words (both with and without a following -n to indicate indefiniteness and definiteness, respectively), rather than vice versa. In this scenario, the whole nominal system would have moved slowly toward the generalization of oblique markers to all contexts, rather than changing one type of marker first, and later changing other types. The timing of these changes may never be known conclusively, but this second possibility deserves to be considered with the more popular first possibility since these early MA documents suggest that it could be plausible.

4.3.4. Accusative -an as an Adverbial Marker

While discussing the occurrences of -an in ASP texts (summarized above in Section 4.3.1), Blau (1966-67: 323-324, 327) notes that adverbs and adverbial constructions which are accusative have the ending -an in ASP more frequently than other types of accusatives do (although they, like all occurrences of -an in these texts, are often omitted as well). Furthermore, in the Greek/Arabic psalm, -an occurs only on an adverb, and in two manuscripts -an is never omitted on adverbs but is omitted other places where it would have been appropriate in CA. Blau concludes that -an had been reinterpreted as an adverbial suffix only (one of its functions in CA) and that non-adverbial occurrences of -an were no longer seen as case markers but were without function. Then the non-adverbial occurrences of -an began to be lost because they had no function, while the adverbial occurrences of -an were retained longer because they had a function. He states, furthermore, that six occurrences of -an on adverbs where CA would have used the nominative or genitive show that -an was extended to these new positions because it was now seen as an invariable adverbial marker and had no function as a case marker.

This conclusion that -an was seen as an invariable adverbial marker in some instances is reasonable, but the evidence does not require that this is the only function that -an had. Since the texts show more frequent instances of -an marking adverbs in positions where it would have occurred in CA than they show -an in any other function it had in CA, it seems that -an as a marker for adverbs was more salient than -an as a marker for any of its other CA functions. If it had high salience as an adverbial marker, it would also have been easy and not surprising for speakers to have extended it to other words that became interpreted as adverbs, as two of Blau's six examples indicate. In these, CA ḥīna ʔiḏ-in-GEN 'at the time when' was interpreted as one word and was spelled with the -an suffix in two manuscripts: ḥīnaʔiḏ-an ACC 'then'. So these examples do not show that -an had taken on a new, invariable function, but only that one of its CA functions was still salient and productive. There is even an indication that this productivity began in late CA, so that ASP was not innovative regarding the productivity but was continuing a trend that had already started. Blau records that in ASP, ʔawwalan often occurs for CA ʔawwala 'formerly', and he

notes (1966 67: 324 n.23) that this is attested in late CA.

Blau's other four examples, while being consistent with the hypothesis that -an was seen as an invariable adverbial marker, also do not require this as the reason for their use. So they, too, do not show a need for Blau's conclusion that -an was seen only as an adverbial marker at the time of ASP. Two of these occur in the manuscript which uses -an in every syntactic position, so the reason for the use of -an on the two adverbs might easily be that the copyist used -an everywhere, rather than that the copyist used -an to mark adverbs. The other two occur after prepositions, as shown below with the prepositions and the occurrences of -an underlined.

ASP:

(1) bi-yaqīn-an
with-certainty-case marker

'certainly'

(2) wa min baʕd qalīl-an
and from after little-case marker

'and not long after'

Therefore, these examples may be evidence that -an had become an invariable adverbial marker regardless of the case that had been required by CA. However, Blau also lists two other examples of -an used incorrectly (according to CA) in the ASP texts following a preposition, when the words it is attached to are not adverbs. These are listed below with the prepositions and -an underlined.

(1) ʔila mawḍuʕ[sic] ʕaḏīm-an wāsiʕ-an jiddan
to place big-case marker wide-case marker very'

'to a big and very wide place'

(2) wa-kān-ū ʔanās min banī ʔasqū yuhūdiy-an
and-was-3MASC.PL.PERF people from tribe Sceva Jew-case marker

'and there were some sons of Sceva, a Jew'.

The existence of these last two uses of -an following a preposition show that sometimes -an was used incorrectly (according to CA) without being an adverb. Therefore, the instances in which it was used incorrectly and was an adverb could have been due to the general reason that CA rules often were not followed, rather than to a specific change of ASP using -an as an invariable adverbial marker.

Therefore, since the uses of -an as an adverbial marker in ASP do not necessarily show--contrary to Blau--that -an was seen in ASP as an invariable adverbial marker, a different explanation may provide a more consistent account of their occurrences. Since, as Blau notes, -an is often missing in ASP manuscripts even when it would have been used in CA to mark adverbs in the accusative, this indicates that the sense of -an as an adverbial marker

was not extremely high in ASP. When considering this with the fact that -an appears in the ASP texts as an accusative adverbial marker more than it occurs to mark other accusatives as it had in CA, the strongest conclusion that can be drawn about -an is that ASP speakers retained a sense of it more as an accusative adverbial marker than as any other type of marker, but that even this function was not extremely salient to them.

Seen in this perspective, then, the use of -an to mark adverbs in ASP is not very different from its use to mark other functions in ASP--contrary to Blau's claim. Therefore, this function, which Blau discusses as an exception to the pattern he proposes, can instead be seen as part of the general pattern proposed above in Section 4.3.2--that ASP was undergoing the process of extending oblique markers (including -an) to all contexts. This would explain the last six of Blau's examples discussed above in which -an was used in non-accusative contexts--whether marking an adverb or not. In fact, this explanation would provide a coherent account of all the facts about adverbial -an, while Blau's account raises the questions discussed above. The adverbial -an data can, therefore, be taken as additional support of the theory proposed here, since they show one more way that this theory provides a coherent account of otherwise somewhat problematic and seemingly unrelated facts.

5.0. Conclusions

This reanalysis of data provided by Blau (1966-67) on the Middle Arabic Southern Palestinian Christian Arabic dialect has suggested a number of new conclusions about the characteristics of nouns in this dialect and the changes that brought about these characteristics. These conclusions support some previous analyses and call others into question. This study has shown the following:

(1) The data is consistent with Blau's and Cantineau's claim that a stress shift occurred, thereby creating a situation favorable for vowels to weaken and drop. However, there is not enough data to confirm this hypothesis.

(2) Contrary to Blau's assertion that case endings had been dropped already, the data show that case endings had only sometimes been dropped at the ends of nouns and before pronoun suffixes. Case endings had sometimes been retained in form in these positions but had ceased to carry out their case marking function.

(3) Contrary to Cantineau and in support of Birkeland and Blau, the data show that loss of the single vowel case endings cannot have been due just to the loss of final short vowels--something must have caused internal changes as well. This could have been either elision of unstressed vowels or generalization of pausal forms to context position.

(4) The data support Blau's and Cantineau's assertion that the accusative case may have been the last case lost in singular and broken plural nouns. It also suggests more than these theories--that the accusative ending was optionally extended to the positions of the other case endings.

(5) The data show that the nominative case may have been replaced with the oblique case throughout the nominal system, not just in dual

nouns and in masculine sound plurals as Blau indicates. Furthermore, generalization of the oblique case may have begun before final case vowels were lost.

(6) As Blau asserts, the data show that accusative -an was retained more consistently as an adverbial marker than in its other functions. However, contrary to Blau, the data indicate that this was not an exceptional phenomenon but that it was part of the pattern of generalizing oblique markers throughout the system (#4 above).

One final point should be made. The change proposed here that the oblique case (which is often considered to be a marked case), rather than the nominative case (which is often considered to be an unmarked case), was generalized throughout the nominal system in Arabic does not follow what has been claimed to be the most usual direction of morphological change--that unmarked forms generally replace marked forms rather than vice versa (cf. Mańczak 1957; and Bybee & Brewer 1980). The situation proposed here is not unknown in changes in case systems, though. For example, the accusative case was the basis upon which the singular paradigm was remade between Ancient and Modern Greek, and it was generalized in the Romance languages as they evolved from Latin. Although a number of different factors influence the direction of morphological change--markedness and frequency being very influential, although not always the most influential (cf. Greenberg 1966, 1969; Mańczak 1957; and Tiersma 1982, who summarizes previous work on markedness and frequency in morphological change and discusses some systematic exceptions), it would be reasonable for the ASP oblique marker--which included the greater number of cases (two)--to be the one that was generalized throughout the system while the nominative marker--which included only one case--was lost.

Notes

*I would like to thank Brian Joseph for his helpful comments on several drafts of this paper.

1. Blau uses these terms frequently in his discussions of MA but does not define them anywhere. The closest he comes to an explanation is to say (Blau 1961: 213) that since CA preserves 'short vowels in open unstressed syllables, it seems necessary to assume a weakly centralizing stress. In Middle Arabic, however, stress has become strongly centralizing, as may be inferred from a large number of phenomena,...: final short vowels have disappeared...; final long ones have been shortened...; and even in the interior of the word short vowels in open unstressed syllables have been elided....' Since Blau uses preservation vs. shortening and disappearance of vowels to define the types of stress, I assume he means that the stress is either weak (which would allow vowels to be preserved) or strong (which would allow vowels to be shortened and dropped), and both types also result in centralization of the vowels. It does not seem to me that he means that the vowel centralization was first weak and then strong.

2. As Blau (1961: 206-207, 220; 1966: 39) notes, Fück (1950: 5, 57-62)

discusses Middle Arabic briefly, but he relies on his intuitions about its origins, saying that the details are nearly unknown because there is no evidence available from that time. The manuscripts analyzed by Blau overcome this problem, since they begin only two generations after the beginning of the Islamic conquests and so provide evidence from essentially the time that Arabic had the impetus from the conquests to change drastically.

3. Blau (1961: 220, 224; 1965: 6-8) states that these early official Muslim papyri were probably written by scribes who were not native Arabs, but that, since these scribes were no doubt from the upper stratum of society, they were probably imitating the speech of their Arab masters and so were reflecting the speech of the Arabs in the documents. He also notes that the few deviations from CA found in these papyri are like the deviations found in Arab poetry and ḥadīth writings of the time. While the first reason is rather speculative, the other two reasons are more conclusive, and so the conclusion seems reasonable.

4. In his summary of Judaeo-Arabic, Blau (1965) specifically declines to take a stand on the role that a change in stress played in the loss of the case endings for this dialect. However, Blau (1965: 168-169) argues that hypothetical phonetic laws and extension of pausal forms to context probably brought about the loss of case endings, with stress playing an important role in some of the dialects.

5. Since the case vowels for these two words when followed by connected complements are the only case vowels that are long in CA, and long vowels are the only vowels that are normally indicated in the Arabic script, the vowels in these examples may also be the only case vowels that are indicated in all of the Arabic ASP writings.

6. Sometimes a change is said to start where an element is redundant, since the element is apparently not needed there because its purpose is also indicated by another element. For example, Corriente (1971, 1973) argues that a cause for the loss of the Arabic case endings was their redundancy. (But see Blau's 1972 reply.) However, elements are also sometimes retained longest in environments where they are redundant--as is claimed here. In Greek, for example, the infinitive is retained longest in contexts in which its subject is uniquely determinable (e.g. after can and begin) and it is, therefore, redundant.

References

Birkeland, Harris. (1952). Growth and structure of the Egyptian Arabic dialect. Oslo: I kommisjon Hos Jacob Dybwad.

Blau, Joshua. (1961). The importance of Middle Arabic dialects for the history of Arabic. Scripta Hierosolymitana IX.206-228.

_____. (1965). The emergence and linguistic background of Judaeo-Arabic: A study in the origins of Middle Arabic. London: Oxford University Press.

_____. (1966-67). A grammar of Christian Arabic, based mainly on South-Palestinian texts from the first millennium, 3 vols. Louvain: Imprimerie Orientaliste.

_____. (1972). On the problem of the synthetic character of Classical Arabic as against Judaeo-Arabic (Middle Arabic). The Jewish Quarterly Review, New Series 63.1.29-38.

Bybee, Joan, and Mary Brewer. (1980). Explanation in morphophonemics: Changes in Provençal and Spanish preterit forms. Lingua 52.201-242.

Cantineau, Jean. (1953). Le dialecte arabe de Palmyre. Bulletin de la Société de linguistique de Paris 49. Paris. [Reprinted from Mémoires de l'institut français de Damas, i, Beyrouth, 1934.]

Corriente, Federico C. (1971). On the functional yield of some synthetic devices in Arabic and Semitic morphology. The Jewish Quarterly Review, New Series 62.1.20-50.

_____. (1973). Again on the functional yield of some synthetic devices in Arabic and Semitic morphology. The Jewish Quarterly Review, New Series 64.2 154-163.

Ferguson, Charles. (1959). The Arabic koine. Language 35.1.616-630.

Fück, Johann. (1955). 'Arabīya. Recherches sur l'histoire de la langue et du style arabe. Trad. par Claude Denizeau. Paris: Librairie Marcel Didier.

Greenberg, Joseph. (1966). Language universals. The Hague: Mouton.

_____. (1969). Some methods of dynamic comparison in linguistics. Substance and structure of language (ed. by Jaan Puhvel), 147-203. Berkeley and Los Angeles: University of California Press.

Mańczak, Witold. (1957). Tendences générals des changements analogiques. Lingua 7.298-325, 387-420.

Tiersma, Peter Meijes. (1982). Local and general markedness. Language 58.4.832-849.

OSU WPL 34 (1986) 84-107

ONE RULE OR MANY? SANSKRIT REDUPLICATION AS FRAGMENTED AFFIXATION*

Richard D. Janda
University of New Mexico
and
Brian D. Joseph
The Ohio State University

0. Overview

Linguistics, it is well-known, is a heavily comparative discipline. For one thing, the simultaneous comparison of various related languages is universally recognized as an absolute necessity in historical reconstruction. Furthermore, though, crosslinguistic typologizing of diverse unrelated languages is also now increasingly accepted as an indispensable step in elaborating even synchronic grammatical theory. There is thus ample justification for beginning a discussion of Sanskrit reduplication and its broader implications by first citing an Armenian joke and then retelling it with an Indic twist.

A whole humorous literature exists of questions allegedly submitted to Radio Yerevan, which broadcasts from the similarly-named capital of the Armenian Soviet Socialist Republic. These queries invariably receive the response "In principle, yes" or "In principle, no", usually followed, though, by additional comments which have the effect of completely retracting the original answer. For example: Question--"Radio Erevan, is it true that Comrade Ivan Ivanovich won a shiny new Volga automobile in the last drawing of the State lottery?"; Answer--"In principle, yes. But it wasn't Ivan Ivanovich; it was Sergei Sergeyevich; and it wasn't a new Volga, it was an old bicycle; and he didn't win it, it was stolen from him while he was inside buying his ticket." Hence, further, along similar lines: Question--Radio Yerevan, is it true that Sanskrit reduplication involves only a single, straightforward rule whose elegant autosegmental treatment as just a special subtype of affixation supports the crosslinguistic validity of such an approach?"; Answer--"In principle, yes. But Sanskrit reduplication isn't one simple rule; it's a collection of many heterogeneous rules, with varying degrees of complexity; and its resemblance to

---

EDITOR'S NOTE: This paper first appeared in ESCOL '85. Proceedings of the Second Eastern States Conference on Linguistics, (Department of Linguistics, The Ohio State University, 1986), pp. 103-119. It is being reprinted here (by permission of the publisher), with some minor changes (mostly corrections, but also four new footnotes, all marked with a) since it is representative of an on-going line of investigation into morphological change that the authors are undertaking, and thus reflects work in historical linguistics at OSU.

nonreduplicative affixation isn't confined to overall formal similarity; this also extends to their parallel morpholexical fragmentation, as in numerous other languages; and, finally, Sanskrit reduplication doesn't always involve copying; sometimes it is so prototypically affixal that it isn't really reduplication at all."

The basic conclusions adumbrated above in the foregoing vein can now (with all due sobriety) be reformulated and summarized as follows:

I. Sanskrit reduplication is not a single rule, but a constellation of several distinct rules.

II. These rules are best analyzed as parallel to those for nonreduplicative affixes.

III. The above analysis is motivated not only by evidence from individual stages of Sanskrit but also by the considerable historical evidence pointing to increased fragmentation of reduplication over time.

IV. Fragmented reduplication--and lexical particularization of morphological rules in general--is not limited to Sanskrit, but appears to be the cross-linguistic norm.

These conclusions are supported by a solid body of general and specific evidence, which we present in the sections below.

1. On Reduplication in General

The overall phenomenon of reduplication has lately been the focus of intense investigation. Building on earlier findings by Wilbur 1971, Moravcsik 1978, and Carrier 1979, recent work by McCarthy 1979, 1981, Marantz 1982, Broselow 1983, Broselow & McCarthy 1983, and numerous others[1], has helped bring this process into the forefront of current research on multilinear phonology and morphology. However, the facts of reduplication in Sanskrit--though well-described and readily available ever since the appearance of Whitney's classic grammar in 1889--have received surprisingly little attention in the aforementioned literature. Similarly, the separate body of contemporary research devoted to Sanskrit linguistics has treated reduplication in that language only tangentially, mainly in connection with discussions of Grassmann's Law and related issues.[2]

Here, however, Sanskrit reduplication constitutes the central focus. Quite apart from the challenge it presents for language-particular description, this phenomenon bears directly on a number of significant issues in morphological theory. These include: (i) the degree of homogeneity shown by reduplication processes in individual languages, (ii) the characterization of reduplication as affixal or not, (iii) the nature and variety of affixal templates, and (iv) the relative value that grammars should place on semantic specificity as opposed to phonological generality in morphological and lexical rules.

The most important theoretical finding to emerge from this study, however, is that Sanskrit reduplication requires the adoption by morphological theory of a new construct, the rule-constellation. This construct can be defined as a group of formally similar morphological

processes sharing at least one characteristic property of form but distinguished by individual formal idiosyncrasies which prevent their being collapsed with one another. This concept is reminiscent, as a formal inverse counterpart, of the functional notion of rule-conspiracy in phonology, and it also recalls the notion of sloppy identity in the syntactico-semantic sphere of anaphora. For Sanskrit, the rule-constellation of reduplication involves several word-formation processes which all indeed have in common a partial prefixal template, but each of which additionally requires a unique set of further template-material and segmental prespecifications.

Equally important, though, is the related finding that the rule-constellation of Sanskrit reduplication reveals a preference on the part of speakers for fragmented morphological and even lexical processes--ones which are semantically specific at the expense of phonological generality, rather than phonologically general at the expense of semantic specificity. Since the same preference for morpholexically particularized word-formation rules is exhibited in numerous languages other than Sanskrit, grammars in general must be constructed so as to reflect this preference.

Such conclusions thus provide a kind of back-handed support for the autosegmental-morphological view of reduplication as affixation that is currently in vogue. On purely formal grounds, there can be little disagreement that reduplication--however analyzed--qualifies as an extreme subtype of context-sensitive morphological addition, and hence as affixation. But the prevalent autosegmental justification for such a treatment is that it obviates the need for transformational formalism and so permits an economical and elegant treatment of reduplication which can easily be assimilated to that of straightforward rules of garden-variety (non-reduplicative) affixation. Based on the evidence from Sanskrit, however, it seems that reduplication and nonreduplicative affixation also show a striking similarity to each other precisely in their apparent inelegance. That is, the morpholexical fragmentation of reduplication in Sanskrit and other languages is exactly paralleled by the existence of fragmented morphological rule-constellations clustering around the unquestioned affixes of, again, Sanskrit and other languages. A most convincing piece of evidence for this parallelism is discussed below (section 5), in which a Sanskrit reduplication-subrule was reanalyzed as a rule affixing an invariant prefix--certainly the ultimate in the intersection of reduplication, fragmentation, and prototypical affixation.

2. Evidence for the distinctness of the Sanskrit reduplication rules

Reduplication is found in a large number of formations in Sanskrit, within both the nominal and the verbal systems. Attention here is focussed on reduplication in the verbal system, where five categories of stems--present, aorist, perfect, desiderative, and intensive--all show reduplication.[3] If it were the case that all such formations involving reduplication behaved alike with respect to a variety of grammatical phenomena, then one would be justified in speaking of "reduplication in Sanskrit" as if it were a unified

process. However, such is not the case, for there are in fact many significant formal differences in the various ways that reduplication manifests itself in the language. Taking note of these differences is the first crucial step in demonstrating that Sanskrit reduplication is indeed a "rule-constellation".

For example, the prosodic template associated with reduplication formations is most commonly CV-, as the underlined elements in (1) show:[4]

1. ta-tap- (perfect stem of tap- 'heat')
   vi-vak- (present stem of vac- 'speak')
   du-dru- (aorist stem of dru- 'run')
   śu-śut-sa- (desiderative stem of śudh- 'purify')

However, there are several other forms that this template can take. Thus, in addition to the CV- type in (1), there is also (treating long vowels (V:) as VV):

2. a. V-: e.g. u-va:c- (perfect stem of vac- 'speak'), i-nak-ṣa- (desiderative of naś- 'attain')[5]
   b. VV-: e.g. a:-v- (perfect stem of av- 'favor')
   c. VC-: e.g. a:n-ams- (perfect stem of aś- 'attain'), ʌm-am-a- (aorist stem of am- 'injure'), al-ar- (intensive stem of ṛ- 'go'),[6] iy-ar- (present stem of ṛ- 'go')
   d. CVV-: e.g. va:-vac- (intensive stem of vac- 'speak'), ja:-jval- (intensive stem of jval- 'burn, flame'), mu:-muc- (aorist stem of muc- 'release')
   e. CVC-: e.g. bad-badh- (intensive stem of ba:dh- 'oppress'), dan-dah- (intensive stem of dah- 'burn')
   f. CVCV-: e.g. kari-kṛ- (intensive stem of kṛ- 'make')
   g. CVCVV-: e.g. gani:-gam- (intensive stem of gam- 'go'), mari:-mar- (intensive stem of mṛ- 'die').

Admittedly, some of these shapes are restricted to particular categories: e.g., CVCV(V)- (as in (2e/f)) is found only in the intensive formation. And there are, to be sure, predominant patterns in any given category (e.g. CV- in the perfect, present, and desiderative), and some evidence of regularization of irregular formations, e.g. the Vedic perfect stem of vap- 'strew' u-va:p- ---> later va-va:p-. However, the existence of these divergent shapes cannot be discounted, and they cannot be assigned to just one type of category; rather within each verbal grammatical category with reduplication, a number of template shapes are to be found. Thus, the evidence of the variety of prosodic templates used in categories with reduplication supports the notion that reduplication in Sanskrit cannot be viewed as a unitary process with a single template valid for all reduplicative formations.

A second feature which differs in the various reduplicative formations is the feature prespecification for the consonant(s) in the reduplication syllable. The predominant pattern is for these consonants to be [-aspirated, -back], so that a velar in the root is

reduplicated as a palatal and an aspirated consonant is reduplicated as a nonaspirated consonant. This pattern is illustrated in (3), and note especially (3e) where it is the second consonant that is reduplicated as [-aspirated]:[7]

3. a. ja gam- (perfect stem of gam- 'go')
   b. bi-bhed- (perfect stem of bhid- 'split')
   c. ji-ghra:- (present stem of ghra:- 'smell')
   d. ji-ga:-sa- (desiderative stem of ga:- 'go')
   e. bad-badh- (intensive stem of ba:dh- 'oppress').

However, there are also a few formations in which [+aspirated] and/or [+back] consonants appear in the reduplication syllable(s). For example, the Vedic subtype of the intensive formation with disyllabic reduplication does not follow the predominant [-aspirated, -back] pattern seen above in (3):[8]

4. a. gani:-gam- (intensive stem of gam- 'go')
   b. bhari:-bhṛ- (intensive stem of bhṛ- 'bear')
   c. ghani:-ghan- (intensive stem of han- 'smite').

Moreover, in later Sanskrit, there is a reduplicative adverbial formation which allows [+aspirated] consonants in the reduplication element, e.g. ratha:rathi 'chariot against chariot' (cf. ratha- 'chariot').[9]

Thus, reduplication syllables do not all reflect a uniform consonantal prespecification in their templates.

Nor is it the case that reduplication syllables follow a uniform template prespecification for vocalism, giving yet another reason for treating the various reduplications in Sanskrit as formally distinct processes. In particular, there is no constant vowel quality or vowel length across all reduplication rules: the examples in (1) through (4) above show a, i, u, a:, u:, and a--i: in the reduplication syllable(s), and other vocalisms are to be found as well:

5. a. bi:-bha:y- (perfect stem of bhi:- 'fear')[10]
   b. ti:-tap- (aorist stem of tap- 'heat')
   c. ne-nij- (intensive stem of nij- 'wash')
   d. bo-bhu:- (intensive stem of bhu:- 'become')[11]

Thus, it clearly is not possible to state a single pattern for the vocalism of the reduplication syllable(s) that is valid across all reduplication types in the language. However, each reduplication rule does have one most common, unmarked value (and a variety of marked values) for the length and quality of the "rhyme" of the reduplication syllable, e.g. i in the desiderative and present, i: in the aorist, a copy of root vowel in perfect, and a heavy reduplicative syllable (e.g. a long vowel or diphthong, though CVC(V(V)) is possible too) in the intensive. The examples in (1) through (5) above illustrate these categorially-based uniformities, as well as some of the marked divergences within each category. Such

a situation alone points to at least five distinct reduplication rules for the verbal system, and the marked subvarieties may well give evidence for the need for further fragmentation in the description of reduplication in Sanskrit.

One additional striking difference in the various reduplications lies in the placement of the reduplication syllable. In particular, the reduplication syllable is mostly prefixed, as in all the examples above,[12] but there is a subclass of desideratives and another of aorists (both formed from vowel-initial roots) in which there is internal reduplication, with a -Ci- reduplication syllable being infixed before the final consonant of the root. A few examples of this small but mildly productive class are given in (6):

6. a. e-di-dh-iṣa- (desiderative stem from edh- 'thrive')
   b. a:-pi-p-a- (aorist stem from a:p- 'obtain')
   c. e-di dh-a- (aorist stem from edh-, cited only in native grammatical literature)
   d. ar-ji h-iṣa- (desiderative stem from arh- 'deserve', cited only in native grammatical literature).

Certain of these forms, especially those with no change in the reduplicated consonant, e.g. a:pipa-, could even be considered to have reduplicative suffixes (i.e. with an analysis [a:p-ip-a-]), though the clear cases (where there is a change in the reduplicated consonant) seem to have internal placement of the reduplication syllable.[12a] While this type probably arose by a reformation of an earlier form with prefixed reduplication (perhaps *id-idh-) to ed-idh- by analogy to the root vocalism[13], this latter form admits of synchronic analysis into a discontinuous root e...dh- with infixed reduplication (-di-). The fact that this pattern was also extended to other such roots suggests that this is the analysis that (at least some) speakers actually made.

A final difference among the various reduplications in the verbal system of Sanskrit concerns certain root idiosyncrasies associated with reduplication. In particular, five roots show a "reversion" of the root-initial palatal to a velar consonant in various reduplication categories, but this reversion is not found uniformly across all the categories for those roots. For example, while the reversion always occurs in the desiderative, it otherwise is scattered across the remaining categories. The following is an (attempt at an) exhaustive listing of the relevant forms, grouped according to root, which show reduplication categories where reversion occurs and, where this can be determined, those where it does not:[14]

7. a. ci- 'note': ci-ke- (present stem), ci-ki: ṣa- (desiderative stem), ci-ka:y- (perfect stem)
   b. cit- 'perceive': ci-ket- (perfect stem), ci-kit-sa- (desiderative stem), ce-kit- (intensive stem), but, cited by native grammarians: ci:-cit- (aorist stem), ci-cet- (alternative perfect stem)[15]

c. ji- 'conquer': ji-ga:y- (perfect stem), ji-gi:-ṣa- (desiderative stem), but ji:-jay- (aorist stem), je-ji:y- (intensive stem, from native grammarians)
d. hi- 'impel': ji-ghy-a- (present stem)[15a], ji-ghi:-ṣa- (desiderative stem, from grammarians), but ji:-hay-a- (aorist stem, from grammarians)
e. han- 'smite': ji:-ghan- (aorist stem), jaṅ-ghan- (intensive stem), ja-ghan- (perfect stem).[16]

Taken together, then, these facts concerning formal differences in the various manifestations of reduplication in Sanskrit point clearly to the conclusion that it is misleading to speak of reduplication in Sanskrit as if it were a unitary process. Instead, a good many reduplication subrules are needed--for observational as well as for descriptive adequacy.

3. Evidence for the Clustering of the Various Sanskrit Reduplication Rules

Despite the conclusion just drawn from the facts in section 2, there are, nonetheless, some striking ways in which the various reduplication rules are formally similar. These facts constitute the second crucial step in demonstrating that the reduplication rules form a rule-constellation, since they show that the rules have some formal properties in common.

The first such property is a trivial one, but must be mentioned nevertheless. As is clear from the examples in (1) through (7) above, all reduplication templates contain at least a vowel. Moreover, in a fully autosegmental analysis, they would all be marked with the feature [+reduplication].

There are, however, more significant common features. In particular, all reduplication rules show the same regular contrast in the differential copying of root consonantism with initial sibilant (ṣ, ś, s, though this last involves a perhaps somewhat nonstandard use of the term sibilant) clusters. Where the second segment in the cluster is a stop, only the stop is copied, i.e. #S(ibilant) + (s)T(op)...- ---> T-vowel-ST...-, but where the second segment is a resonant, the sibilant is copied, i.e. #S + R(esonant)...- ---> S-vowel-SR...-. Examples of the stop-type are given in (8a); examples of the resonant-type, in (8b):

8. a. sthā:- 'stand' ---> ti-ṣṭha- (present stem)
spṛś- 'touch' ---> pa-spṛś- (perfect stem), pi-spṛk-ṣa- (desiderative stem)
stu- 'praise' ---> tu-ṣṭav- (aorist stem)
sthi:v- 'spew' ---> te-ṣṭhi:v-/te-ṣṭhi:v- (intensive stems, from grammarians)
ścut- 'drip' ---> cu-ścot- (perfect stem), cu-ścut-a- (aorist stem)
b. śru- 'hear' ---> śi-śra:v- (perfect stem), śu-śruv- (aorist stem)
smṛ- 'remember' ---> sa:-smṛ- (intensive stem), su-smu:r-ṣa- (desiderative stem), both from grammarians.

This is the one significant formal feature common to all reduplications in Sanskrit without any exceptions.[16a] It is important to stress "formal" here, for it is the case that in general, reduplication is not unitary within categories from a functional and/or semantic standpoint. With the exception of the intensives, whose stems always have reduplication of some sort, there are nonreduplicative formations to be found in each of the categories that show reduplication.[17] By its unique commonality to reduplication, this feature gives some unity to what is otherwise, from a formal standpoint, an assortment of numerous different rules.[17a] Still, given the rather large number of features on which the reduplication rules disagree, it seems best to conclude that they do indeed form a rule constellation, united primarily in the way they treat sibilant clusters and related in the fact that they all involve at least a vowel prefix, but distinct nonetheless in their behavior with respect to a wide variety of formal aspects.

4. Contrast with Other Analyses with Little or no Recognition of Fragmentation

While Sanskrit reduplication has been mentioned quite extensively in the generative literature (see footnote 2), the view of Sanskrit reduplication taken here is an entirely novel one. One notable exception is Schindler 1976, which talks (p. 627) of the remnants of Grassmann's Law in Sanskrit as "one of several morphological rules that apply ... [in] redu- plication".[18] For the most part, previous researchers in the generative framework have either acted as if the reduplication rules in Sanskrit were a unitary process, or as if there were at least unity within categories. Thus, Sag (1976, p. 617) gives "the reduplication rule" as:

9. $[_{\text{ROOT}}\ \underset{1}{\begin{bmatrix}\text{C}\\ 2\\ \text{velar}\end{bmatrix}}\ \underset{3}{\text{V}}\ \underset{4}{\text{X}}] \rightarrow \underset{2}{\begin{bmatrix}\text{palatal}\\ \text{asp}\\ \text{etc.}\end{bmatrix}}\ 3\ 1\ 2\ 4.$

Similarly, Cairns & Feinstein (1982, p. 210-1), following Kiparsky (1979, p. 434-5) declare that "the Sanskrit [reduplication] rule will have the form __copy Mc " [= margin core of root syllable's onset], and Marantz (1982, p. 448-9n.9) speaks simply of "Sanskrit initial reduplication". Anderson (1982, p. 602), on the other hand, implicitly recognizes categorial differences in reduplication, but nonetheless gives a single rule for "the reduplication in ... ["the perfect stem"]":

10. $\begin{bmatrix}\text{+Verb}\\ \text{+Perfect}\end{bmatrix} / \left(\begin{bmatrix}\text{+coronal}\\ \text{+cont}\end{bmatrix}_{1} \right) \begin{bmatrix}\text{syllabic}\\ \text{+obstruent}\end{bmatrix}_{2} \text{C}_0\ (\text{a})\ [\text{+syllabic}]\ \text{X}\ /$

$__\ /\ 2\ 5\ 1 \qquad\qquad 2\ 3\ 4\ 5\ 6$.

Similarly, Borowsky & Mester (1983, p. 53) refer to the "the formation of the perfect [in Sanskrit] ... by prefixing a template CV- to the root and copying and associating the segmental melody", though recognizing some categorial differences by referring (p. 61n.2) to "[some] intensive forms [which] involve reduplication of the entire root morpheme". To a certain extent, these analysts were simply giving the unmarked or predominant formative process in each case, but such oversimplifications dangerously obscure the actual quite fragmented picture. As noted already (and in the next section), this fragmentation is to be expected, given the affixal nature of reduplication, so the contrast here is not just one of detail but rather one of substance.

5. Parallels Between Reduplicative and Nonreduplicative Sanskrit Affixes

Calling reduplication affixal in nature means treating it as not distinct in any meaningful way from the (nonreduplicative) affixes of the language. This view has been argued for by Marantz 1982 for reduplication in human language in general, and it certainly holds for Sanskrit, based on both synchronic and diachronic facts which show that a number of parallels obtain between (undisputed) affixes and reduplication in Sanskrit.

From a synchronic standpoint, there is considerable motivation for treating reduplicative elements as a type of affixation. Most importantly, doing so fills out holes in the distribution of both reduplication and (nonreduplicative) affixation. That is, while there are numerous (nonreduplicative) suffixes in Sanskrit, there is but one grammatical prefix, the past tense marker a- (the so-called "augment", see footnote 12). Similarly, while the placement of the reduplication element is mainly prefixal, one subpattern of the desiderative and aorist subtypes noted above in (6)--namely those forms that show no change in the reduplicated consonant (e.g. a-pip-a-) -admits of analysis as having reduplicative suffixes. Thus by treating reduplication as a type of affixation, the one otherwise anomalous grammatical prefix, the augment, ceases to an irregularity, and the one type of anomalous suffixal reduplication likewise is no longer irregular. Moreover, it can be noted that there are both reduplicative infixes, as in the type of (6) with changes in the reduplicated consonant (e.g. e-di-dh-), and nonreduplicative infixes, such as the formative -na-/-n- which forms the present stem of some 29 roots, including those in (11):[19]

11. a. yuj- 'join- ---> present stem yu-na-j- ("strong")/yu-n-j-("weak")
    b. rudh- 'obstruct' ---> present stem ru-na-dh- ("strong")/ru-n-dh-("weak")
    c. chid- 'cut off' --- > present stem chi-na-d- ("strong")/chi-n-d-("weak").

In terms of the distribution of their placement with respect to roots, then, an affixal treatment of reduplication serves to eliminate irregularities both in reduplication and in the (undisputed) affixes.

From a diachronic standpoint , though, the evidence is even stronger, for many of the things that have happened to affixes in Sanskrit have also happened to reduplication syllables. This parallel behavior suggests that speakers treated the two alike.

For example, both affixes and reduplication syllables sporadically underwent a loss of their identity due to their reanalysis as part of a root. Thus the synchronic root Cpinv- 'fatten' (evident in, for instance, the perfect stem pi-pinv-) represents a reanalysis of an earlier present stem from the root pi(:)- 'swell, fatten' formed with the often factitive suffix -nu-, i.e. *pi-nu-. Similarly, the originally reduplicated intensive stem ja:-gr- 'wake' was reanalyzed as an indivisible root ja:gr-, which is evident, for instance, in the 1SG present ja:gr-mi,[20] and in nominal derivatives such as ja:gar-aka- 'waking'.

Another development found with both affixes and reduplications involves the obscuring of original boundaries and distributions, in what may be called accretions and extensions. Typically, these happen by some type of reanalysis. Thus, the locatival adverbial suffix -ta:t, added more or less pleonastically to other adverbials, e.g. pra:k-ta:t 'from the east' (cf. pra:ñc- 'forward, east'), puras-ta:t 'before; in/from the east' (cf. puras 'in front, forward'), was resegmented to -sta:t, presumably in forms such as purasta:t, and then extended to other forms, e.g. upari-sta:t '(from) above' (cf. upari 'above'). Similarly, the "union"-vowel i/i: was originally part of roots (due to the Indic treatment of Proto-Indo-European root-final laryngeals) but came to be considered part of adjoining suffixes, creating virtual allomorphs of the suffixes, so that the agentive -tr- gained the allomorph -itṛ-, the desiderative -sa- gained the allomorph -iṣa-, etc. In somewhat parallel fashion, the reduplicative intensive prefix with CVR- shape that regularly occurred only with roots containing a resonant was extended, with an -n- that was originally proper only to roots with a nasal, to other roots, e.g. jaṅ-gah- (intensive of gah 'plunge', and cf. the alternative intensive stem with no final -C- in reduplication, ja:-gah-, cited only in the grammarians). Moreover, at some point in the development of the intensive reduplicative prefixes involving reduplication of the whole root, an i:, of somewhat uncertain origin, accreted onto the reduplicative prefix, giving forms such as bhari:-bhṛ- (from bhṛ- 'bear'), and ultimately becoming part of a disyllabic subpattern for intensives (see also footnote 8).

Most significant, though, for the view advocated here is the fact that, in at least one instance, a reduplication syllable, even though its connection to the root was reasonably transparent, was reanalyzed as an affix: a:n-aṃś-, the perfect stem of aś- 'attain', and a:n-aṃj-, perfect stem of aṃj- 'anoint', served, via the identification of the a:n- as merely an affix, as the basis for highly anomalous perfect stems of other roots with initial a- or ṛ-, e.g. a:n-ṛdh-, perfect stem to ṛdh- 'thrive', a:n-ṛh-, perfect stem to arh- 'deserve', etc. The fact that a reduplication syllable could move so

easily to an existence as an affix suggests again that reduplication is in actuality a type of affixation.

Given these facts about the parallel diachronic behavior of reduplication syllables and affixes, and given the synchronic motivation for treating them in parallel fashion, one can legitimately question even calling the phenomenon "reduplication" in all instances. Especially in the case of the a:n-rdh-/ a:n-rh- type of perfect (cf. above and (6)) and of such irregularities as perfect stem ja-bhar- (versus expected and later-occurring ba-bhar-) from bhr- 'bear', there seems to be no reason to speak of "reduplication" except in order to stress a parallel with other formations in the same category. Thus, some context-sensitive prefixation probably is not reduplication at all. Moreover, if one speaks instead of "affixes" in these and the other cases, then a:n-rdh- (etc.) can be said to contain a perfect affix that happens to have no direct formal connection with the root it attaches to, while a more regular formation such as ta-tap- contains one that does have such a direct formal connection.

Sanskrit reduplication, then, is not only best treated as a fragmented constellation of related morphological processes, but further, these processes are best taken as affixal in nature.

6. Diachronic Fragmentation of Morphological Rules in Sanskrit

The rule constellation of Sanskrit reduplication not only is synchronically fragmented but also can be shown to have arisen via the diachronic fragmentation of an originally more unitary situation. This conclusion rests on a mass of philological evidence that can only be summarized here.

Sanskrit is unquestionably a historical development of Proto-Indo-European (PIE), and the source of Sanskrit reduplication is ultimately to be sought in this proto-language. Parallels for each of the Sanskrit reduplicated categories are to be found in other Indo-European languages. As a result, the proto-language is standardly reconstructed (as in Meillet 1964) as having virtually all the reduplicated (verbal) categories found in Sanskrit. However, the standard reconstruction (Meillet, pp. 179-182) also shows greater unity within each of these proto-categories with reduplication than is found in Sanskrit; for example the vocalism in the perfect's reduplication syllable is *e, while that for the present is *i, etc. (compare (5)). Similarly, the highly particularized forms such as u-va:c- (as in (2a)), a:n-rdh- (see section 5), ja-bhar- (see section 5), and others, are not reconstructible as such for PIE. Thus, in the development of Sanskrit from PIE, a diachronic fragmentation of reduplication occurred.

Moreover, Sanskrit is attested over a long enough period of time that it not only has a previous history but also an internal one. And, within Sanskrit, idiosyncratic forms such as u-va:c-, a:n-rdh-, ja-bhar-, etc. can be explained only as particularized replacements for more regular forms: u-va:c- through the lexicalization of a

former sound change grown opaque (one deleting *v- before u), a:n rdh via reanalysis and analogy (see section 5), and ja-bhar- probably via contamination. Furthermore, the fragmentation of the rules for nonreduplicative affixes in Sanskrit (as already noted in section 5), can also be shown to be a historical innovation. Thus the Sanskrit-internal evidence of the development of reduplicative and nonreduplicative affixes shows speakers to have exhibited a preference, in many cases, for fragmented morphological rules and processes--i.e. for rule-constellations.

Actually, though, the motivation for this conclusion can be shown to be much stronger and even more compelling, once the perspective is widened to include more languages than just Sanskrit.

7. Morphological Fragmentation as the Crosslinguistic Norm

If the extreme morpholexical particularization of reduplication found in Sanskrit were a completely isolated case, one could perhaps attempt to counter the apparent need for a morphological construct like the rule-constellation by claiming that the Sanskrit phenomenon in question represents merely an accidental and/or highly marked situation. But fragmented reduplication is in fact found in so many languages that any such line of resistance clearly is totally untenable. In every language known to us which utilizes reduplication to mark either a single grammatically-central morphological category or else several morphological categories (whether central or more peripheral), this functional importance and/or variety is always accompanied by at least some degree of morpholexically-particularized formal fragmentation. Thus, reduplication seems to be a rule-constellation, not only in Sanskrit, but also in Kfhehe (Odden & Odden 1985), Madurese (Stevens 1986), Tagalog (Carrier(-Duncan) 1979, 1984), and many other languages too numerous to discuss or even list here. Furthermore, even in languages where reduplication plays a rather minor role (in terms of functional variety and centrality), there is still usually a considerable amount of formal differentiation, as shown for instance by the contrast in English reduplicative (or at least reduplicativoid) forms such as higgledy piggledy versus flim-flam versus din-din, etc.

In order both to cement the crosslinguistic validity of this overall point and to stress that it is not always clearly brought out in the literature, the treatment of a particular language can be cited, taken from one of the most influential recent articles on the importance of reduplication for morphological theory (Marantz 1982, pp. 474-475). After first introducing "Tagalog reduplication" as if it were a single general phenomenon, the discussion then mentions that Tagalog really has at least "three different sorts of reduplication". Finally, a footnote reveals that, even though the analysis sometimes proceeds "as if the various reduplication prefixes ... are each single, uniform morphemes ... [--a]ctually, ... each prefix has a variety of uses ... [so that] each must be understood as the morphological form of a set of homophonous morphemes." Here, too, then, the notion of reduplication as a rule-constellation is arguably present, implicitly lurking just below the surface.

Moreover, synchronic and diacharonic fragmentation of nonreduplicative affixation--in fact of morphological processes in general--is also extremely common crosslinguistically. Two straightforward cases from English can be adduced involving adjectival suffixes. The alternation between -al and -ar was once phonologically conditioned, but recent pairs such as line-al/line-ar and famili-al/famili-ar show that the -al/-ar contrast no longer represents allomorphy, but instead two separate nearly-homophonous suffixes clustering as an affixational rule-constellation around the formal core Vowel + Liquid. Likewise, such disparate forms as drink-able, pot-able, comfort-able, and surviv-able (as used of nuclear weapons which are--unfortunately--intended to survive, rather than be survived) demonstrate that there are now several homophonous affixes -able. That is, the different morphosyntactic conditions embodied in their various formal statements prevent them from being collapsed with one another (see also Aronoff 1976). German similarly has a two-element rule-constellation for adjectival suffixes expressing material composition (e.g. Seide/seid-en 'silk'/'silken', with -(e)n, versus Stein/stein-ern 'stone'/'stony', with -ern, where final -n is shared), and at least a three-part rule-constellation for agentive suffixes (e.g. Dien-er 'server, servant' versus Tisch-ler 'table-maker, cabinet-maker' versus Red-ner 'speaker, orator'--where final -er is shared).

Nor are nonaffixational cases of morphological rule-fragmentation hard to come by. For example, a phonologically rather arbitrary set of tone-substitution processes in Copala Trique (see Janda 1982a, Hollenbach 1984, and references there) performs the three functions of deriving adjectives from nouns, inflecting nouns for possessedness, and inflecting relative clause verbs for continuative aspect. In addition, a subtractive process of final vowel deletion in Rotuman marks the "incomplete phase" (see e.g. Janda 1983b, 1984, and Hoeksema & Janda 1985), but this category turns out to be merely a convenient cover-term for a set of uncollapsible distinctions including indefinite nouns, verbs in the imperfective aspect, emphatic words, and nonfinal elements in a noun phrase. In fact, the same farrago of categories are all arguably sometimes marked by morphological metathesis in Rotuman, too, and a similar process of permutation is involved in a rule-constellation of Clallam (see the references noted above for Rotuman). Surely the most extensive (and hence most fragmented) rule-constellation currently known, though, is instantiated by Modern High German Umlaut (see Janda 1982a, 1982b, 1983a), which has been morpholexically particularized so severely that it not only occurs alone in six different rules, but also occurs with thirty distinct phonological shapes of affixes, which themselves represent at least twice that many morphological rules. In total, then, the Umlaut constellation demonstrably involves between sixty and seventy morphological rules, most of which share an identical formal core, but some of which are strikingly different in their structural descriptions.

In light of the evidence just presented, it can thus be said, without exaggeration, not only that fragmented reduplication is not

limited to Sanskrit, but also that morpholexical fragmentation--of reduplication, of nonreduplicative affixation, and of morphological processes in general--is indeed the crosslinguistic norm, both as a synchronic state and as a diachronic change. Such states and changes must be interpreted to reflect a strong and constant tendency on the part of speakers to particularize (formerly) more general morphological processes as markers of more specific lexical and grammatical categories. That is, given the notion of the morpholexical rule-constellation as a way to express the unity of similar rules even in the face of their formal diversity, what emerges as the dominant and driving force in creating such constellations is the lexico-semantic motivation of speakers: the high value that they seem to place on the unambiguous and even redundant transmission of information about specific meanings as expressed by particular lexical items (mainly morphemes and words) and classes of lexical items. For example, the occurrence of the usual German agentive marker -er is far less revealing about what stem precedes it than that of its co-members in the constellation (its "co-stars") -ner and -ler, precisely because the latter have a much more limited distribution (and also express certain connotations which -er lacks).

It appears that the morpholexical fragmentation at issue here is fed by three main sources (although a full discussion of such topics must be deferred until a later time and place). First, there is morphologization (and lexicalization) of formerly purely-phonological processes, which often transfers the conditioning for such a process from a single phonological configuration to several morphemes which once had something to do with that configuration. In this way, a once-unitary formal operation can become fragmented via its multiple direct association with numerous affixes and/or roots (German Umlaut being a notorious case in point). Second, there is accretion by metanalysis, whereby a reanalysis of morpheme boundaries results in the effective addition to an existing morpheme of segments which formerly belonged to another morpheme (as illustrated by the German agentives mentioned above, of which -ner and -ler are the result of accretions to -er based on resegmentation of forms like Rechn-er 'calculator, and Regl-er 'regulator', respectively).

Third and finally, but perhaps most commonly, there is reanalysis of root + affix combinations in such a way that not segments but rather semantic and/or morpho(phono)logical properties of a particular root or roots are reassigned (or jointly assigned) to the affix, which thereby becomes correspondingly fragmented from other instances of the formerly identical affix occurring with different roots. It is apparently in this way that Sanskrit forms such as the aforementioned a:n-ṛdh- arose: the reanalysis of words like a:n-aṃś- as having an invariant initial morpheme a:n- rather than a reduplicative affix made available a prefix a:n- which could then be used elsewhere. This mechanism can perhaps be most vividly expressed by the following metaphor: when a given affix is deposited in the bank of the lexicon along with a particular amount of principal contained in a specific root, the account draws semantic and morpho(phono)logical interest mainly on the root or the entire word,

but such interest can be taken along when the affix in question is withdrawn for use with another root.

The mechanisms just described as conspiring to produce morphological and lexical fragmentation (i.e. rule constellations) can be characterized as operating diachronically, but such a characterization by no means absolves linguistic theory of the responsibility to account for such phenomena. In fact, it does just the opposite, given the usual generativist assumption that language change is governed primarily (if not exclusively) by constraints of synchronic grammar. Actually, then, the evidence presented herein regarding speakers' preference for fragmented--that is, morpholexically particularized, rule-constellational--analyses of reduplication and morphological processes in general requires that grammars be constructed so as to place a premium on morpholexical solutions to linguistic problems. Generative grammar, having--correctly, it seems--made a diachronic bed governed primarily by synchronic principles, is here forced to lie in it: the historical frequency and ubiquity of morpholexical fragmentation leave one little choice but to build not just a place but even a preference for such rule-"mitosis" and the resulting rule-constellations into (synchronic) grammatical theory. Moreover, the explanatory potential of such an approach is extraordinarily great, for it promises to cover not only the fragmentation of morphological rules and its associated morphologization of phonological rules, but also "downgrading", the morphologization of syntactic phenomena brought about by such interacting processes as semantic bleaching, cliticization, and clitic-to-affix conversion. Still, limits of space preclude a fuller discussion of such issues at this point, so a number of observations are presented in conclusion regarding general lessons for the elaboration of morphological theory that emerge from this particular study of fragmented reduplication in Sanskrit.

For one thing, there can be no substitute for fine-grained studies of particular instances of a given phenomenon (e.g. reduplication) in a single language (e.g. Sanskrit) as the prime source for revealing insights into the nature of such a crosslinguistically common morphological process (again, as reduplication). Studies which superficially draw only selected "representative" data from a wide range of languages, ignoring exceptional forms as uninteresting and focussing on elegantly-describable forms, simply will not do. They proceed a little like the drunk who dropped his keys in the dark just outside the tavern-door but went looking for them up the street under the lamppost because the light was better there. Actually, they are even less defensible than this, methodologically, because they tend not to take a fair look even for what they seek, but rather to start out with an artificially limited preconception of what they will consider as relevant data. It is not surprising then that reduplication looks crosslinguistically elegant under an autosegmental analysis if in fact the primary data going into such an analysis have been selectively gathered from languages so as to favor straightforward reduplication rules while passing over exceptional and complicated rules of this type.[21]

Next, a related issue, it must be concluded that the attempt to exclude such exceptional and complicated forms from consideration on the grounds that they are somehow "marked" is not only circular but also completely undercut by the fact that fragmented reduplication rules (and rule-constellations in general) are staggeringly comon both across and within languages. When the situation of having many allegedly marked phenomena (like fragmented reduplication) is the unmarked case in languages, then markedness itself is probably best excluded from consideration as an explanatory factor in attempts to account for such fragmentation, or at least needs to be set aside for urgent reevaluation.

Penultimately, the essence of morpholexical particularization of rules as an activity that is lexico semantically driven should bring to mind that the study of formal aspects of word formation does not exhaust the subject matter of morphology. Rather, even what appear to be purely formal characteristics of word formation, such as fragmented reduplication in Sanskrit and many other languages, may often turn out to have some lexico semantic motivation. One should keep in mind here Jakobson's dictum that language without meaning is meaningless.

Finally, the entire phenomenon of morpholexical rule fragmentation and rule-constellations bears directly on issues raised by the frequently heard saying (apparently due originally to Meillet (1903-1904, p. 641) that "le langage forme un système ... où tout se tient" ('language forms a system where everything holds together'). Whether implicitly or explicitly, this claim is surprisingly often taken to mean that all aspects of linguistic structure are equally directly and equally closely linked to one another, and this interpretation then results in a principle that those linguistic analyses are always to be preferred which yield a maximum of structural homogeneity and interconnectedness. While laudable in the abstract, such a view tends ultimately to have a Procrustean effect, since it encourages the brute force ironing out of recalcitrant details within one domain of linguistic analysis on the basis, not of internal considerations, but of fit with other areas of grammar. Moreover, hard evidence against such a practice is available from such research as Ferguson's work on simplified registers like foreigner-talk, which in certain crucial respects appear to be independent of the rest of grammar. As Ferguson (1981, p. 3) boldly put it: "Tout does not se tenir", i.e. though everything in language obviously holds together with something, it is not the case that everything holds together with everything.

Such a conclusion could hardly be more strongly supported by the facts of fragmented Sanskrit reduplication and of crosslinguistic morpholexical particularization in general. As such phenomena evolve, they involve gradual but steady formal and functional development of morphological and lexical rules away from each other, with no apparent regard for anything else in grammar except the expression of lexical and lexical class semantics (and of grammatical-morphemic notions). In fact, the historical linguistic literature is replete with similar instances where locally motivated

changes in grammar led to complexity (and often subsequent change) elsewhere. Given the difficulty which even full-time professional linguists have in keeping the entire grammar of a single language in mind at once, is it actually at all surprising that naive speakers behave as if they are unable to do this? Rather, it seems that their conscious and unconscious dealings with language are severely constrained in scope by a highly limited window determining how much grammatical structure they can consider at one time. At the very least, positing such a limitation seems the appropriate step to take in accounting for fragmentation of reduplication and other morphological processes in Sanskrit and elsewhere. In this way also, such notions as Wittgenstein's family-resemblance and Rosch's prototypes, especially as they have already been brought into linguistics by other scholars, suggest themselves as having much explanatory promise.

An approach like this has a bright future, then, but it also has an estimable past. Although the phenomena of Sanskrit fragmented reduplication led us to the concept and name "rule-constellation" prior to our encountering relevant work by Louis Hjelmslev, that linguistic pioneer turns out to have anticipated such a notion in a general way nearly half a century ago, and by way of conclusion, his words on this overall topic are given (in our own translation, from the French of Hjelmslev 1939/1959, p. 114): "The famous maxim according to which tout se tient dans le système d'une langue ['everything holds together in the system of a language'] has frequently been applied in too rigid, too mechanical, and too absolute a fashion. One must keep matters in proper proportion. It is important to recognize that everything holds together, but that everything does not hold together to the same extent, and that alongside interdependencies, there are also purely unilateral dependencies, as well as pure constellations." It is our hope that, in the present study of Sanskrit reduplication as fragmented affixation and of its broader implications, we have given such ideas as those just quoted a concrete enough form so that they not only can receive a principled answer from Radio Yerevan but will also find practical application in the morphological and general-linguistic investigations of other scholars.

Notes

*We would like to thank Lloyd Anderson, Joan Bybee, Donna Gerdts, John Joseph, Michael Silverstein, Alan Stevens, and Sarah Grey Thomason, for helpful comments made at ESCOL 85. Wolfgang Dressler and George Cardona have made several extremely useful comments on an earlier version of the paper. We are also grateful to Mary Clark, Alan Hudson-Edwards, Ann Miller, Joel Nevis, and Arnold Zwicky for other kinds of help in the preparation of this work.

1. Besides unpublished papers, the numerous other works include Yip 1982, Bell 1983, Carrier-Duncan 1984, ter Mors 1984, Everett & Seki 1985, and Odden & Odden 1985.

2. Stemberger 1980 has a review of recent generative literature on this topic; Borowsky & Mester 1983 is the most recent proposal (evaluated in Joseph & Janda (in preparation)).

3. We do, however, refer on occasion to facts about nominal reduplication. We exclude from consideration here more sporadic types of reduplication such as the a:mredita compounds, consisting of a repetition of a word, with loss of accent in the second member, for an "intensive, or a distributive, or a repetitional meaning" (Whitney section 1260), e.g. vayám-vayam 'our very selves' (cf. vayám 'we'). Still, clearly any comprehensive treatment of the variety of reduplication rules in Sanskrit would have to take such types into consideration as well.

4. We are admittedly being somewhat eclectic in our choice of examples, taking them from all chronological stages of Sanskrit (e.g. Vedic as well as Classical Sanskrit). Our primary source is Whitney (1885, 1889), two of the classic Western grammatical studies of Sanskrit. In part our eclecticism stems from our belief that knowledge of the older language persisted into at least the beginnings of the Classical period (e.g. Vedic forms are noted in Pa:nini's grammar) and so was an area of at least passive competence for many speakers. Also, many patterns we present here as illustrating a certain type can be found in a variety of stages of the language, even if a particular example may be restricted to one period. We have not in general indicated the age of any given form, except where such information is important.

5. The identification of this form is complicated a bit by the existence of another form of the root, aś-, the desiderative of which follows the formation with nonprefixal reduplication discussed in (6).

6. These last three forms are from the Vedas and Bra:hmanas only; there appear to be no VC- nonperfects to be found in Classical period.

7. It may be, though, that the lack of aspiration on the d in bad-badh- is the result not of prespecification of [-aspirated] for the entire reduplication syllable, but instead of the independent workings of Bartholomae's Law (giving an intermediate stage /badbhadh-/ from underlying /badh-badh-/) and Grassmann's Law (giving the attested form).

8. This pattern is found in later stages of Sanskrit as well, but by Classical times, the predominant [-aspirated, back] prespecification prevails, as in Classical bari:-bhṛ- (intensive stem of bhṛ- 'bear', cf. (4b)) and jari:-hṛ- (intensive stem of hṛ- 'take', cf. (4a, c)). Note also that we are here purposely excluding forms such as the desiderative dhitsa- from dha:- 'put' (i.e. /dhi-dh-sa-/) which show aspiration probably as a result of analogy and which are synchronically formed--despite their diachronic origin in reduplication--via an internal change process limited to (a subtype of) desideratives--see Sag 1976 and Schindler 1976 for some discussion of these desideratives as well as other such forms with aspiration in the apparent reduplication syllable. However, bringing

in such forms could only bolster our claims about the lack of uniformity in consonantal prespecification.

9. That this is not simply a dvandva (copulative) or a:mredita (distributional) compound but instead a true adverbial derivation via reduplication is shown by the fact that there is a fixed pattern of vocalism for the unit--the first element always ends in -a:- and the second element in -i. See footnote 3 above, though, concerning the compound reduplications.

10. The form with the vocalism -i:- is found in the rather late Vedic A:itareya A:ranyaka, with -i- found both earlier and later on.

11. If surface e/o in Sanskrit are to be treated as underlying diphthongs (i.e. respectively, as ay/aw, with consonantal y/w), then (5c, d) may not really illustrate differential vocalism.

12. When the aorist stems are used to form a true past tense, an inflectional prefix a- is added outside of the reduplication syllable, e.g. a-ti:-tap-at 'she heated' (aorist of tap-). Similarly, lexical prefixes can be added outside of the reduplication syllable.

12a. George Cardona (personal communication--1/25/86) has suggested that, following Pa:nini, one might analyze the edidhisa type of reduplicated stem as involving rightward iteration (copying) of the second syllable of the root-plus-desiderative suffix sequence (i.e. edhisa- ---> edhidhisa- ---> edidhisa-) rather than the infixation into the root itself of a reduplication syllable, as we suggest here. To a certain extent, the analyses are not really very different, for in both of them, the reduplication syllable is infixed--in our analysis, it is infixed in the root, while in the other it is infixed within the stem formed by the addition of the desiderative suffix. Nonetheless, we believe that a Pa:ninian-style analysis is to be rejected, for two reasons. First, the intermediate stage /edhidhisa-/, as is evident above, must become [edidhisa-] by a deaspiration process. This deaspiration seems in all respects to be similar to the deaspiration (the remnant of Grassmann's Law) that is regular in reduplication when two aspirated consonants come to occur in successive syllables (see above in section 2 and the forms in (3)). However, the edidhisa-type deaspiration must be triggered by a nonroot segment, since the second aspirate is the copied (reduplicated) element, whereas normally (with the exception of the clear relic forms jahi 'strike!' (2SG.IMPV of han-), bodhi 'become!' (2SG.IMPV of bhu:-), and vidatha 'distribution' (derived from dha:- 'distribute')--see Schindler (1976: 626)), only segments that are part of the root trigger the deaspiration. Thus the Pa:ninian analysis requires a complication in the statement of the deaspiration process. Second, not only is there deaspiration (as regularly in reduplicated syllables, in our analysis) in the edidhisa-type of formation, but there is also palatalization of back consonants, as shown by arjihisa-, desiderative stem of arh- 'deserve'; as noted in section 2 (and see the forms in (3)), such palatalization also is regular with segments in reduplicated syllables. In the Pa:ninian analysis, there is no reason to expect an intermediate stage

/arhi-hi sa / to develop into the attested arjihiṣa , since hi is a permissible sequence phonotactically, and the leftmost hi is part of the root not part of the reduplicated syllable. However, under our analysis, the ji in arjihiṣa is the reduplication syllable, so that the palatalization is expected. Thus, our root infixing analysis presents no complications in the phonological processes associated with reduplication.

13. Admittedly, it is quite difficult to know exactly what the expected forms should be, since vowel-initial roots present a rather mixed assortment of reduplications of all types, even at the earliest stages; the pattern of edidhiṣa- is a likely candidate to have been the model since it is attested somewhat early (in the late Vedic Va:jasaneyi-Samhita:) and since a plausible path of development can be inferred for it. Indeed, the limited spread of the edidhiṣa- type suggests that this provided for speakers a relatively satisfactory solution to the problems posed by these vowel initial roots.

14. We omit here one form, the Vedic hapax legomenon jáguri- 'steep (?), leading (?)', not only because of its obscurity, but because of disagreement as to its etymology. Only if it is connected with jṛ 'waste away', as Whitney (1885: 55) believes, does it show reversion. Mayrhofer (1956), however, more plausibly connects it with gur- 'lift up', citing phonological problems with the jṛ- etymology (since jṛ- is from Indo-European *g'er-, and the palatal *g' should never yield a Sanskrit [g]). We are also excluding sporadic instances of reversion not connected to reduplication, such as Rig Vedic 3Pl middle aorist a-sṛg-ran from sṛj- 'send forth'.

15. The native grammarians divide these forms into two roots, kit- and cit- , despite their etymological identity, so that under such an analysis, there is no reversion to speak of, but only parallel formations from parallel roots.

15a. As George Cardona has kindly pointed out to us, this form, despite Whitney's (1885: 205) clasification of it as connected to hi- 'impel', may in fact represent a different root.

16. If the difficult epic apparent intensive 2SG imperative jí:jahi is a form of han- , then this would be an example from this root without the reversion.

16a. It is important to point out, as Wolfgang Dressler has kindly reminded us, that the occurrence of T...ST in the reduplications of the (8a) type is not the result of a general phonotactic constraint operative in the language prohibiting sequences of ST...ST in successive syllables or within the same word. Forms such as astoṣṭa '(s)he praised' (3SG s-aorist of stu- 'praise') and stha-sṭa '(s)he stood' (3SG sa- aorist of stha:- 'stand') show that the situation found in these reduplications is not a matter of Sanskrit phonotactics, and thus the pattern illustrated in (8a) is probative for demonstrating the clustering of the various reduplication rules.

17. Thus, there are many nonreduplicated present stems (ten classes in all, with reduplicated stems making up only one class, representing 4% of the total--to go by Whitney's statistics) and aorist stems (seven formations in all, of which reduplicated stems make up only one type, representing 21% of the total--again based on Whitney's statistics), a subclass of desideratives without reduplication (see footnote 8 above), and even a few perfect stems with no reduplication (e.g. ved- from vid- 'know').

17a. It might be thought that the generalization illustrated by the forms in (8a) is a static truth about the language but not necessarily one that shows that speakers actually made the connection among the various rules in the way we suggest they did. However, two of the sibilants that participate in the (8a)-type reduplication are sounds that developed within the history of Indo-Iranian and/or Indic (the palatal ś being the Indo-Iranian and Indic outcome of the Proto-Indo-European palatal stop *k' and the retroflex ṣ being a specifically Indic development of *s in a variety of environments). Thus the fact that all three sibilants behave alike with respect to reduplication of *ST- roots shows that speakers at some point in the history of Sanskrit made the generalization across reduplication rules that we are claiming, thus providing a degree of unity for the various rules in question.

18. It is, moreover, the view that is implicit in most of the traditional grammatical studies of Sanskrit, e.g. Whitney 1889.

19. This is the native grammarians' seventh class. "Strong" versus "weak" forms of the stem are distributed according to morphological category, e.g. singular versus plural, active versus middle, etc. The underlying /n/ of the infix changes to [n̥], [n̄], etc. in predictable phonological environments.

20. As a reduplicated present, ja:gṛmi would be unusual in having a weak grade second syllable, and the absence of any intensive meaning is noteworthy.

21. In fact, another detailed study of Sanskrit reduplication and related topics has recently appeared (Steriade 1985), although we gained access to it too late to permit further account to be taken of it in this paper. Still, it is significant that Steriade's paper, while involving a different focus and approach from ours, completely supports the idea that Sanskrit has far more than just one reduplication rule, and so provides independent motivation for our proposed concept of "rule-constellation". Steriade employs a Marantzian autosegmental approach to reduplication in Sanskrit, and this gives us a chance to add here one final remark on this general type of analysis. We firmly believe that the essentials of the autosegmental view of phonology and morphology have much to contribute to the analysis of reduplication both in Sanskrit and crosslinguistically, and our analysis in this paper is couched mainly in autosegmental terms. Nevertheless, we are not convinced of the necessity of adopting the Marantzian variation on this theme whereby the entire phonemic melody of a root or stem is copied over a

reduplicative template, regardless of how much of that melody actually appears on the surface. For arguments in favor of alternative approaches involving more limited copying (albeit with freer copying power), see Janda 1984, Hoeksema & Janda 1985.

References

Anderson, S. R. 1982. "Where's Morphology?". LIn 13.571-612.

Aronoff, M. 1976. Word Formation in Generative Grammar. MIT Press.

Bell, S. J. 1983. "Internal C Reduplication in Shuswap", LIn 14.332-338.

Borowsky, T. & R. A. Mester. 1983. "Aspiration to Roots. Remarks on the Sanskrit Diaspirates". CLS 19:52-63.

Broselow, E. I. 1983. "Salish Double Reduplications: Subjacency in Morphology", Natural Language and Linguistic Theory 1.317-346.

Broselow, E. I. & J. McCarthy. 1983. "A Theory of Internal Reduplication", Linguistic Review 3.25-88.

Cairns, C. E. & M. H. Feinstein. 1982. "Markedness and the Theory of Syllable Structure". LIn 13.193-225.

Carrier[Duncan], J. L. 1979. The Interaction of Morphological and Phonological Rules in Tagalog. MIT Ph.D. Dissertation.

_____. 1984. "Some Problems with Prosodic Accounts of Reduplication". In M. Aronoff and R. Oehrle, eds., Language Sound Structure. Cambridge: MIT Press, pp. 260-286.

Everett, D. & L. Seki. 1985. "Reduplication and CV Skeleta in Kamaiurá", LIn 16.326-330.

Ferguson, C. 1981. "Simplified Registers and Linguistic Theory", unpublished paper, Stanford University (prefinal draft of (revised) paper published in L. K. Obler & L. Menn (eds.), Exceptional Language and Linguistics. New York: Academic Press, 1982).

Hjelmslev, L. 1939/1959. "La structure morphologique", Vme Congrès International des Linguists. Rapports. Bruges, 1939, pp. 66-93 (reprinted in revised form in Essais linguistiques. Travaux du Cercle linguistique de Copenhague 12 (1959):113-138).

Hoeksema, J. & R. D. Janda. 1985. "Implications of Process Morphology for Categorial Grammar". Paper presented at the Conference on Categorial Grammars & Natural Language Structures, Tucson, AZ, May 31-June 2.

Hollenbach, B. E. 1984. The Phonology and Morphology of Tone and Laryngeals in Copala Trique. University of Arizona Ph. D. Dissertation.

Janda, R. D. 1982a. "Of Formal Identity and Rule-(Un)-Collapsibility: On Lost and Found Generalizations in Morphology". In D. Flickinger et al., eds., WCCFL 1:179-197.

-----. 1982b. "High German Umlaut and Morphologization", unpublished University of Arizona ms.

-----. 1983a. "Two Umlaut-Heresies and Their Claim to Orthodoxy". University of Arizona Coyote Papers 4:59-71.

-----. 1983b. "'Morphemes' Aren't Something That Grows on Trees: Morphology as More the Phonology Than the Syntax of Words". In Richardson et al., eds., Interplay. Chicago: Chicago Linguistic Society, pp. 79-95.

-----. 1984. "Why Morphological Metathesis Rules Are Rare: On the Possibility of Historical Explanation in Linguistics". BLS 10:87-103.

Joseph, B. & R. Janda. (In Preparation). "Sanskrit as She is Misanalyzed Autosegmentally".

Kiparsky, P. 1979. "Metrical Structure Assignment is Cyclic". LIn 10.421-441.

Marantz, A. 1982. "Re Reduplication". LIn 13.435-482.

Mayrhofer, M. 1956. Kurzgefasstes etymologisches Wörterbuch des Altindischen. Heidelberg.

McCarthy, J. 1979. Formal Problems in Semitic Phonology and Morphology. MIT Ph. D. Dissertation.

-----. 1981. "A Prosodic Theory of Nonconcatenative Morphology", LIn 12.373-418.

Meillet, A. 1903 1904. "[Review of] Michel Bréal. Essai de sémantique. Science des significations, 3e édition revue, corrigée et augmentée, Paris, 1904", L'Année sociologique 8:640-641.

-----. 1964. Introduction à l'étude comparative des langues indo-européennes. University, Alabama: University of Alabama Press (reprint of 8th edition of 1937).

Moravcsik, E. A. 1978. "Reduplicative Constructions", in J. Greenberg (ed.), Universals of Human Language. Vol. 3: Word Structure. Stanford: Stanford Univ. Press, pp. 297-334.

Odden, D. & M. Odden. 1985. "Ordered Reduplication in Kihehe". LIn 16.497-503.

Sag, I. 1976. "Pseudosolutions to the Pseudoparadox: Sanskrit Diaspirates Revisited", LIn 7.609-622.

Schindler, J. 1976. "Diachronic and Synchronic Remarks on Bartholomae's and Grassmann's Laws", LIn 7.622-637.

Stemberger, J. P. 1980. "Another Look at Grassmann's Law", Glossa 14.113-135.

Steriade, D. 1985. "Reduplication, Ablaut and Syllabicity in Sanskrit", unpublished MIT ms.

Stevens, A. 1986. "Reduplication in Madurese", ESCOL '85. Proceedings of the Second Eastern States Conference on Linguistics Columbus, OH: Department of Linguistics, The Ohio State University, pp. 232-242.

ter Mors, C. 1984. "Affix to X: Broselow & McCarthy Versus Marantz: A Reinterpretation", Linguistic Review 3.275-298.

Yip, M. 1982. "Reduplication and C-V Skeleta in Chinese Secret Languages", LIn 13.637-61.

Whitney, W. D. 1885. The Roots, Verb Forms, and Primary Derivatives of the Sanskrit Language. Leipzig: Breitkopf und Härtel.

_____. 1889. Sanskrit Grammar[2]. Cambridge: Harvard University Press.

Wilbur, R. 1973. The Phonology of Reduplication. University of Illinois Ph. D. Dissertation.

OSU WPL 34 (1986) 108-122

Fragmentation of Strong Verb Ablaut in Old English*

Keith Johnson

1. Introduction

It has recently been claimed (Janda & Joseph, 1986; Bybee, 1986) that the normal line of development for morphological processes of language is from simple to complex, from unitary processes to fragmented ones. This claim goes completely counter to the standard generative approach to language change in which change is viewed as a type of simplification (Kiparsky, 1968). Janda & Joseph cite several examples of fragmentation (among them are Sanskrit reduplication, the English adjectival suffixes *-al, -ar,* and *-able*, and German umlaut) which lead them to the conclusion that 'morpholexical fragmentation - of reduplication, of nonreduplicative affixation, and of morphological processes in general - is indeed the crosslinguistic norm, both as a synchronic state and as a diachronic change' (Janda & Joseph, p. 113). In their view and in the view expressed by Bybee (1986) fragmentation as a diachronic development is the normal, natural line of development for morphological processes.

In this paper I demonstrate another case of morphological fragmentation - strong verb ablaut in Old English (OE). The conclusion of the paper is that the process of ablaut which was inherited by OE from Indo-European (IE) was fragmented as the result of sound changes and that this fragmentation resulted in particularization of ablaut patterns to smaller and smaller sets of lexical items. One of the theoretical implications of this account is that sound change seems to normally work by changing lexical representations, not by adding, deleting or simplifying phonological rules. Also, the facts reported here seem to indicate that the synchronic result of fragmentation (in this case) is a pattern of relations within the lexicon not a set of rules (i.e. not a rule-constellation[1]). Finally, the study indicates that speakers/hearers focus on the signifying functions of specific lexical items rather than general structural patterns (Bybee's (1986) independent development vs. structural coherence)[2].

2. Strong Verb Ablaut in OE

Verbs in Germanic can be classed as weak or strong according to whether their past tense is characterized by an alternation in the stem vowel. Traditionally, the strong verbs are divided into seven classes according to the structure of their stems. The number and type of post-vocalic segments in the Germanic stem determined the ablaut pattern of the stem vowel. Typical examples of the strong verbs of classes I - VI are listed in table 1.

Sound changes have obscured the pattern of the post-root segments. The first two classes had the vowels *i* and *u*,

respectively, after the root vowel. Class III was characterized by a post-root sequence of a resonant followed by an obstruent. Class IV had a single resonant following the root vowel, while class V had a single obstruent in this position.

Table 1
The OE strong verb, classes I - VI
from Lass and Anderson (1975, p. 25)

Class	Pres.	Pret. Sg.	Pret. Pl.	Part.	
I	bīdan	bād	bidon	-biden	'wait'
II	beodan	bēad	budon	-boden	'command'
III(a)	helpan	healp	hulpon	-holpen	'help'
(b)	weorpan	wearp	wurpon	-worpen	'throw'
(c)	bindan	band	bundon	-bunden	'bind'
IV	beran	bær	bǣron	-boren	'bear'
V	sprecan	spræc	sprǣcon	-sprecen	'speak'
VI	bacan	bōc	bōcon	-bacen	'bake'

2.1 Origin of the ablaut patterns

In this section I will discuss the IE origins of the OE ablaut patterns. I will show that the patterns illustrated in table 1 are not derived historically from the same IE pattern, but constitute a mixture of several types of IE verb forms. This account is taken primarily from Prokosch (1939) and Campbell (1959).

The present tense of the Gmc. strong verbs of classes I - V was characterized by the vowel *e*. This vocalism comes down to OE unchanged in the class IV and V forms *beran* and *sprecan*. The class I sequence *ei* became *ī* in Gmc. The class II sequence *eu* was changed in OE to *ēo*. Two regular sound changes affect the vowel in class III. Breaking produces the diphthong *eo* before *r* in *weorpan*, and raising before nasals resulted in *i* in *bindan*. This latter change also affected some verbs in class IV (eg. *niman* 'to take').

The preterite singular in classes I - V is a continuation of IE perfect (cf. Grk λ(ε)λοιπα, λ(ε)λειπα)[3] and has the IE ablaut vowel **o* > Gmc. *a*. The ablaut forms of the first five classes are thus: (I) **ai* > *ā* (Campbell, 1959, 134); (II) *au* > *ēa* [æ:a] (Campbell, 1959, § 135); (III) **a* > *æ* (Campbell, § 131) except before nasals, then *æLC* > *eaLC* ([æa]) (Campbell, § 143-144); (IV, V) **a* > *æ* (Campbell, § 131).

The preterite plural[4] in the first three classes of strong verbs had reduced grade (hence IE *i,u,R* > Gmc *i,u,uR*), while classes IV and V have Gmc *ē*[5] from IE *ē*. In OE this developed as *ǣ* except before nasals where it became *ō*. The 2sg ending in W.Gmc. was *-iz* > OE *-e*. This ending is probably from an IE

aorist of the type of Grk. ἔ-λιπ-ε. Since some IE aorist formations were also characterized by reduced grade ablaut it is possible that the OE pret. 2sg. is the later development of an IE aorist rather than an IE perfect. An IE aorist origin is also likely for the pret. pl. endings, of which Prokosch (1939) says, 'the Gmc. 3pl. in *-un* cannot be perfect endings, and the 1pl. in *-um* need not be one; the 2pl. in *-uþ* is surely analogical' (p. 163). (OE generalized the 3pl. ending to all pl. forms.) Thus, it is possible that the reduced grade of the vowel in the pret. pl. of classes I - III reflects an IE aorist rather than an IE perfect. These considerations led Prokosh to attribute the pret. pl. of classes IV and V in *ē* to an IE aorist which he also sees in Grk. ἔβην and Lat. *vēn-* and *sēd-* (§ 57). Because these Grk. and Lat. forms have other, more probable origins it is perhaps better to consider the *ē* form as a later development and peculiar to Gmc., although it must have been present at a very early stage of development (all of the Gmc. languages show a reflex of it in class IV & V verbs).

The Michels-Streitburg theory[6] of the origin of a long vowel in classes IV and V is dismissed by Campbell (1959) - 'This theory involves many difficulties (especially with regard to the simplification of consonant groups) and is better abandoned' (p. 305, n. 1). Another objection to this theory is that it provides no principled explanation for the preservation of reduplication in classes IV and V while the reduplicated syllables in the other classes were simply haplologized.

Lass and Anderson (1975) proposed that the long vowels result from a grammatically conditioned rule such as (1) which

(1) Pret. pl. lengthening: from Lass and Anderson (1975)

$$V \rightarrow \left\{ \begin{array}{l} [-long] \\ [+long] / \underline{\quad}\ \overset{}{\underset{Pret2}{}}\ CV \end{array} \right\}$$

lengthens pret. pl. vowels when they are followed by only one consonant (i.e. classes IV and V in Gmc.). This requires that underlying representations be quite abstract. The arguments below in favor of surfacy representations also constitute arguments against this view of the origin of the long vowel in classes IV and V pret. pl. forms.

Wherever the pret. pl. in classes IV and V originated these considerations indicate that 'the Gmc. strong preterite is not a homogeneous development from one source, but a combination of several types' (Prokosch, 1939, p. 164). This is one indication that the strong verbs in OE started out fragmented to some degree. As will be indicated below classes VI and VII present further complications.

The passive participle[7] in the first three classes had the same form as the pret. pl. in Gmc. In classes IV and V instead of being completely reduced (as in I - III) the root vowel was 'schwa

secundum'. In Gmc. this IE sound became *u* before resonant consonants and *e* before obstruents. Thus, pass. part. forms had *u* in class IV and *e* in class V. The *u* which was present in the pass. part. stems of classes II, III, and IV was changed to *o* presumably due to the influence of the vowel in the pass. part. suffix.[8] The sound change was that *u* > *o* before mid and low vowels. 'There are, however, many exceptions in OE' (Campbell, p. 43). The fact that there are very few, if any, exceptions among the pass. part. forms of strong verbs will be discussed below.

For the purposes of this paper it is enough to quote Campbell concerning classes VI and VII. 'The verbs of Classes VI and VII are of a different type' (p. 305). The IE origins of the forms for these classes are not very clear at all (see Prokosch, §§60 - 62 for a

Table 2
Origin of the ablaut patterns: summary.

	pres.	pret. sg.	pret. pl.	pass. part.
Classes:				
I	*e-i>ī	*o-i>ai>ā	*ø-i>i	*ø-i>i
II	*e-u>ēo	*o-u>au>ēa	*ø-u>u	*ø-u>u>o
III	*e>e/_lO >eo/_rO >i/_NO	*o>a>ea/_LO >a/_NO	*øL>L>uL *øN>N>uN	*øL>L>uL>oL *øN>N>uN
IV	*e>e >i/_N	*o>a>æ >o/_N[1]	*ē>ǣ >ō/_N	*$_b$R>uR>oR
V	*e>e	*o>a>ǣ	*ē>ǣ	*$_i$O>eO
VI	*a	*ō	*ō	*a
Endings				
Sg. 1	*-ōm>-ǣ>-e	*-a>-ø		
2	*-isi>-st		*-es>-iz>-e	
3	*-i$_b$i>-t	*-e>-ø		
pl.	*-an$_b$i>-a		*-un >-un>-on	
part.				*-ono>-on> -an>-ǣn>-en
prefix				ge-

[1] *ō* by analogy.

O = obstruent consonants
L = liquid consonants
N = nasal consonants

discussion of the posibilities). Class VI verbs have two likely IE sources (1) thematic presents with suffix accent (also called 'aorist presents') and (2) 'weak presents' with a j-suffix as is common with weak verbs in OE. Class VII includes some vestiges of IE reduplication along with a number of verbs of uncertain etymology. The formal characteristics of the class VI and VII verbs include: (1) the pres. stem and the pass. part. normally have the same vowel, while the pret. sg. and pl. have the same vowel, (2) class VI has *ō* in the pret. sg. and pl., (3) class VII has either *ē* or *ēo* in the pret. sg. and pl. (4) the vowel of class VI pres. and pass. part. is *a* for a large number of cases but there is more variety here than in the pret. forms, (5) class VII pres. and pass. part. forms are most often in *ā, ā, ēa*.

Table 2 is a summary of this section which illustrates the point that OE ablaut was never a unified phenomenon. The Gmc. background of the strong verb forms was a mixture of several different patterns from IE. So, from the very beginning of OE there was not just one ablaut pattern but several.[9]

2.2 Further fragmentation within the classes

This section discusses further types of fragmentation which occur in the OE strong verbs. In table 3 which is a summary of this section 28 different ablaut patterns are listed.

Contract verbs in class I formed an alternate ablaut pattern to the usual one. When Gmc. [χ] occurred intervocalically it was deleted in OE. Thus, from the Gmc. present tense form **wrīhan* 'cover' (h=[χ]) OE has *wrēon* as the result of regular sound changes. This resulted in the ablaut pattern: *ēo ē i i*. Other class I verbs which have this pattern are: *lēon* 'grant', *þēon* 'thrive', *sēon* 'sieve', and *tēon* 'accuse'.

There is a fairly large set of verbs (13 out of 51 basic class II verbs in Borden, 1982) in class II which show a present in *ū* instead of the expected *ēo*. These verbs probably originally had a reduced grade present stem (Prokosch, § 58,b calls these aorist presents). Campbell suggests that the lengthening of *u* may have been by analogy with class I 'verbs with *ai* in the past had *ī* in the present system, those with *au* in the past might develop *ū* in the present system' (§ 736,b). At any rate, another ablaut pattern for class II verbs exists.

Class III is already represented by three different ablaut patterns according to the resonant following the stem vowel. To complicate matters a little more we also note that there are some aorist presents which are normally included in this class also (*spurnan*, 'spurn', *spearn, spurnon, spornen*). Also, there is a set of verbs which were originally class V (with one post-root obstruent) but 'their presents were extended by the addition of a dental element, e.g. OE *streg-d-an* strew, *feoh-t-an* fight,

Table 3

Ablaut patterns in OE strong verbs,
from Levin (1964).

I	1a	ī	ā	i	i	bīdan	bād	bidon	biden	'await'
I	b	ēo	ā	i	i	wrēon	wrāh	wrigon	wrigen	'cover'
II	2a	ēo	ēa	u	o	bēodan	bēad	budon	boden	'command'
II	b	ū	ēa	u	o	brūcan	brēac	brucon	brocen	'use'
III	3a	i	a	u	u	bindan	band	bundon	bunden	'bind'
III	b	e	ea	u	o	helpan	healp	hulpon	holpen	'help'
III	c	eo	ea	u	o	weorpan	wearp	wurpon	worpen	'throw'
III	d	u	ea	u	o	spurnan	spearn	spurnon	spornen	'spurn'
III	e	e	æ	u	o	stregdan	strægd	strugdon	strogden	'strew'
IV	4a	e	æ	ǣ	o	beran	bær	bǣron	boren	'bear'
V	b	e	æ	ǣ	e	metan	mæt	mǣton	meten	'measure'
V	c	ēo	ea	æ	e	seon	seah	sǣgon	segen	'see'
V	d	i	æ	ǣ	e	biddan	bæd	bǣdon	beden	'pray'
VI	5a	a	ō	ō	a	faran	fōr	fōron	faren	'go'
VI	b	ēa	ō	ō	a	slean	slōg	slōgon	slagen	'strike'
VI	c	e	ō	ō	a	hebban	hōf	hōfon	hafen	'raise'
VI	d	ie	ō	ō	ea	scieppan	scōp	scōpon	sceapen	'create'
VI	e	æ	ō	ō	a	st ppan	stōp	stōpon	stapen	'step'
IV	f	u	ō	ō	u	cuman	cōm	cōmon	cumen	'come'
IV	g	i	ō	ō	u	niman	nōm	nōmon	numen	'take'
VII	6a	ā	ē	ē	ā	hatan	hēt	hēton	hāten	'call'
VII	b	æ	ē	ē	æ	lætan	lēt	lēton	læten	'let'
VII	c	ō	ē	ē	a	fōn	fēng	fēngon	fangen	'seize'
VII	7a	a	ēo	ēo	a	bannan	bēonn	bēonnon	bannen	'summon'
VII	b	ea	ēo	ēo	ea	fealdan	fēold	fēoldon	fealden	'fold'
VII	c	ā	ēo	ēo	ā	blāwan	blēow	blēowon	blāwen	'blow'
VII	d	ēa	ēo	ēo	ēa	bēatan	bēot	bēoton	bēaten	'beat'
VII	e	o	ēo	ēo	ō	blotan	blēot	blēoton	blōten	'sacrifice'

Traditional class identifications are given in the first column, followed by the reclassification suggested by Levin based on the preterite vocalism.

frig-n-an ask' (Campbell, § 736,b). Thus, new patterns are added for *stregdan (e,æ,u,o)*, and *frignan (i,æ,u,o)*.

The pattern *i,ō,ō,u* for the class IV verb *niman* 'take' can be derived from Table 2 (although not specifically discussed in the previous section). The aorist present verb *cuman* 'come' has almost the same pattern: *u,ō,ō,u*. Thus, two more ablaut patterns must be added to those already identified for class IV.

There are some contract verbs in class V which result in a different ablaut pattern (*sēon*, see, *seah, sāwon, segen* or in Angl. *sēon, seah, sāwon, sewen*). There is also a set of weak presents which have present forms resembling class I weak verbs. Otherwise they have ablaut forms like normal class V verbs (*biddan*, 'ask, pray', *bæd, bǣdon, beden*).

These same two types of variation in ablaut are found in class VI verbs. Contract verbs such as *slēan*, 'slay' have a pattern different from the normal class VI ablaut pattern (*slēan*, slay, *slōg, slōgon, slagen*). There are also weak presents of three different types -- *hebban*, 'raise' has the pattern: *e,ō,ō,a;* *scieppan*, 'create' has the pattern: *ie,ō,ō,a;*[10] and *stæppan* 'step' has the pattern: *æ,ō,ō,æ.*[11]

Class VII verbs are divided into two basic categories according to the vocalism of the pret. forms (*ē* and *ēo*). There are three different patterns with pret. forms in *ē*, and five different patterns with pret. in *ēo*.

Table 3 is a collection of the different ablaut patterns which have been discussed in this section and the previous one. Levin (1964) chose to class the verbs by their pret. forms and so the traditional classes are spread out to some degree among Levin's classes. The degree to which speakers of OE made such identifications is unknown. The issue will be addressed briefly in section 4 below.

3. Lexicographical evidence concerning the strong verbs

In a survey of the strong verbs listed in Borden (1982) and their corresponding forms in Middle English (Stratmann and Bradley, 1891)[12] I found that they were productively used in word formation processes in OE and that they survived into ME at uniformly high rates.

Table 4 summarizes the results of this survey. The first row indicates the total number of verbs in each class listed in Borden(1982). The second row shows the number of these verbs which are not the result of derivational processes (ie. *helpan* 'help' vs. *āhelpan* 'support'). The percentage of the total which are underived forms (third row) indicates that although their absolute values are quite different all of the classes have roughly the same ratio of basic forms to derivations. These percentages indicate that there are approximately four derivations for every basic form in the strong verbs. This is evidence for a certain degree of productivity.

Table 4

Lexicographic evidence concerning OE strong verbs.

	I	II	III	IV	V	VI	VII	explanation
1.	283	270	423	100	171	192	290	Tot. # of verbs
2.	61	52	85	15	30	27	55	# of basic forms
3.	22%	19%	20%	15%	18%	14%	19%	% basic forms of tot.
4.	8	14	12	4	8	7	17	# of anom. basic fms
5.	13%	27%	14%	27%	27%	26%	31%	% anom. of basic fms
6.	42	40	60	11	22	21	42	# of basic fms in ME
7.	69%	77%	71%	73%	73%	78%	76%	% ME of basic fms

The fourth row in table 4 is the number of basic forms (as listed in the dictionary - ie. present tense form) which do not match the forms predicted by table 2. The percentage (row 5) of the total number of basic forms (row 2) which are anomalous in this sense indicates the degree of fragmentation. With the exception of classes I and III the percentage of each class which does not fit the basic description of the class is about 28%. This is an indication that the fragmentation discussed in section 2.2 is more than just an occasional exception to an otherwise overwhelmingly stable pattern.

Row six is the number of basic forms which came down to ME as strong verbs (as found in Stratmann and Bradley, 1891). In all cases the percentage of OE basic forms which survive into ME is quite high (row 7, $\bar{x}$=74%). This indicates that the fragmentation of OE strong verbs did not cause speakers to avoid using them. They do not seem to be particularly difficult for speakers.[13]

This survey of lexicographical evidence indicates then that although strong verbs were indeed fragmented to a substantial degree they were nevertheless included in word formation processes and were not abandoned over the period from OE to ME.

4. Evidence against abstractness

If the lexical representations of strong verbs were sufficiently abstract then the fragmentation which was demonstrated in section 2 would indicate nothing more about OE than that there were a number of interesting synchronic rules which caused the surface forms of strong verbs to have some variety. It is my claim in this paper that fragmentation resulted in variation in the lexical representations of strong verbs -- not just variation in their surface forms. In order to establish the claim that fragmentation results in changes in lexical representations it is necessary to show that the lexical representations involved are not abstract. There is a wealth of

evidence in the OE strong verbs which rules out abstract representations. Some of this evidence will be presented in this section.

4.1 Analogies

Lexical representations are the representations in memory of the words that speakers/hearers use. They are memories, and like other memories they have varying degrees of strength.[14] When our memory for a word is weak we may rely on the pattern of a similar word which is more clearly remembered.[15] This is analogy. The point is that analogy is from one lexical representation to another in this sense of 'lexical representation'.

There are some cases of analogy in the OE strong verbs (to be discussed below) in which surface forms such as those listed in table 3 are the basis for analogy. This indicates that the lexical representations for these forms are like their surface forms.

The analogical lengthening of the aorist present in class II (table 3, 2b) verbs in Gmc. has already been mentioned. Crucial to this analogy is that the class I present form is *ī* in the lexicon (*ai:ī::au:X*).

Also, the lengthening of *o* in pret. sg. in class IV verbs with post-vocalic nasals (table 3, 5f & g) has been attributed to analogy (in the note to table 2). There are two possible models for this analogy. If the class IV pret. pl. is the model this would indicate that (1) the sound change *ẽ* > *õ*/_*N* was a change in lexical representations, and (2) that the pret. pl. was a lexical representation - not derived form the pres. form. If the class VI pret. sg. served as the model for this analogy the lexical status of the pret. sg. form is indicated.

A variant which occurs beside *nõm* pret. sg. of *niman* (table 3, 5g) is *nam*. This variant 'is due to analogy of nasal verbs of class III' (Campbell, p. 313, n. 1). The formula: *bindan:band::niman:X* indicates once again that the pret. sg. (*band*) is stored in the lexicon.

There is a tendency for contract verbs of class I in *ẽo* to shift to class II. Thus, *tẽon* 'accuse' has *tẽah, tugon* in addition to *tãh, tigon*. About four percent of the class I basic forms shifted to class II in this way. This analogy illustrates the sufficiency of surface identity (*ẽo, ẽo*) for the occurrence of analogy. It also illustrates the structural coherence of the ablaut pattern. The pattern is productive in the sense that it can be extended to new verbs in the class. Note also that it is the surface pattern that is extended.

Finally, the possibility of analogical forces involved in the form of the pass. part. in classes II,III, and IV should be mentioned. As was mentioned earlier the sound change lowering *u*

before a following mid or low vowel was somewhat erratic in application. (Note the exceptions: *ufan* 'over', *ufor* 'higher', *pusa* 'bag', *sugga* 'a kind of fish', and *tube* 'trumpet'.) The fact that it was not eratic in the case of the pass. part. of strong verbs may be an indication of analogy. The *o* vocalism may have taken on some grammatical/semantic value that helped the sound change take place in all pass. part. forms.[16] Once again, separate lexical representations (and non-abstract ones) are suggested by the evidence.

4.2 Non-application of sound changes

Breaking did not apply to *berstan* 'burst' or *þerscan* 'thresh'. These words were originally class V verbs formed by the addition of an infix with a dental consonant (*stregdan* was given earlier as an example). They have also undergone a metathesis. Thus, the Gmc. forms were **bres-t-* and *þre-sk-*.

The non-application of breaking could be taken quite simply as an indication of rule ordering. The derivations would be something like (2).

(2)	/bres-t-an/	/þre-sk-an/
breaking	----	----
metathesis	berstan	þerscan

Campbell (§ 459,1) has metathesis in this case as *rV* > *Vr/_s* or *n*. Exceptions include: *hrespan* 'to strip, spoil', *cranic* 'record', *cranoc* 'crane', *cristalla* 'crystal', *bresne* 'mighty', *brasian* 'to do work in brass', *brastlian* 'to roar, rustle', *restan* 'to rest', *rendan* 'to rend', *scrind* 'swift course', *strand* 'sea-shore', *þrosm* 'smoke', *trandende* 'precipitous, steep', *trendon* 'to turn around', *þrines* 'Trinity', *þrintan* 'to swell', *þryscan* 'to wiegh down'.

The order of the 'rules' in (2) has a phonetically motivated 'low level' rule followed by a lexically specific 'high level' rule.[17] A simpler theory of phonology is possible if we treat breaking as a sound change which changed the lexical representations of the words to which it applied (ie. */werpon/* > */weorpan/* 'throw'). After this sound change had ceased to be active another sound change (metathesis) produced lexical representations which could have undergone breaking if they had existed when breaking was active.

Metathesis also affected the class III verb *brinnan* 'burn' which with metathesis had the forms: *birnan, barn, burnon, burnen* (Campbell, § 741 - compare table 3, 3a). In this case we expect breaking before *r* (like *weorpan*, table 3, 3c) but instead the ablaut pattern does not change at all. In addition to the non-application of breaking, the preservation of the effects of raising before a nasal consonant is interesting. This rule is

formalized by Lass & Anderson (1975, p. 29) as (3).

(3) $$V \begin{matrix} \\ -low \end{matrix} \text{ --> } +high \text{ / } __ +nasal$$

Thus, *bindan* and *brinnan* are derived from *bendan* and *brennan* respectively. The rule also blocks the application of back umlaut in the pass. part. forms *bunden* and *brunnen*. The stages of development for *birnan* were thus:

(4)

Gmc	bren-
raising	brin-
metathesis	birn-

As in the case of the interaction of breaking and metathesis, the historical order of these changes is not the preferred synchronic order. If the lexically specific 'higher level' rule (metathesis) is ordered before the 'lower level' phonetically motivated rule (raising), then metathesis would bleed raising, giving *bernan* instead of *birnan*. It seems preferable in this case to view raising before nasals as a sound change which had run its course before metathesis took place. In this view, (3) is not a synchronic process, but represents the lexicalized results of a sound change.[18] This is more clear evidence against abstract lexical representations.

4.3 Borrowing a Latin verb

Finally, the borrowing of the Lat. verb *scrībere* as a class I strong verb *scrīfan* 'decree' is evidence for the non-abstract representation of class I verbs. Here, as in several of the cases of analogy mentioned earlier, surface identity is enough to associate the verb with class I verbs with present stem vowel *ī*.[19] The borrowing also indicates the synchronic salience of the ablaut pattern, although the fact that this was the only verb borrowed into the strong verb system (while many were borrowed as weak verbs) is evidence of the tendency to generalize the weak verb pattern at the expense of the strong verb pattern.

5. Conclusion

Thus, the synchronic result of the fragmentation described in section 2 existed in the lexical representations of the words involved, not in morphological or phonological processes in the synchronic grammar of OE. The evidence presented in this paper is consistent with a view of language in which the lexicon (as words stored in memory) takes a central position. Although there is evidence (some of it presented in this paper) for the reality and potential for active use of patterns within the lexicon which might be described using process notation, the conclusion to which the evidence points is that these 'processes' are actually patterns in the lexicon. The appearance of behavior which indicates that some sort of association has been made (for instance, the analogy of class I verbs like *wrēon* with class II verbs like *bēodan*) cannot be taken as evidence for the existence of a morphological process. (In fact the evidence in this particular case indicates that a pattern

among some lexical items has been extended to some other lexical entries.)

The overall view of language sound systems which is indicated by the evidence in this paper is that language is rich in the lexicon and poor in the grammar - that most of the information needed to pronounce words is available in their representations in memory and that the modifications which they undergo between lexical retrieval and articulation are quite limited. Along with this view (and also indicated by the evidence presented here) is a view of sound change in which the lexicon is central. Sound change is seen as primarily change in lexical representions rather than grammar change. This view helps to explain why fragmentation occurs and why it is a common, even normal, situation: for each particular sound change some lexical representations have the appropriate conditions for change while others don't; therefore fragmentation can result from a series of sound changes.

Notes

[*]Many thanks to Brian Joseph. He is the one who told me that the only thing my paper needed was to be written, and his knowledge and eye for linguistic detail has saved me from some embarrassing mistakes. Some of my fellow students at Ohio State have also provided useful comments and valuable discussion - Brad Getz, Peter Lasersohn, Joyce Powers, Jane Smirniotopoulous and Debbie Stollenwerk. Thanks also to the Crusaders for musical assistance.

[1]Janda and Joseph (1986) found that the synchronic result of fragmentation in Sanskrit reduplication was a rule-constellation, i.e. 'a group of formally similar morphological processes sharing at least one characteristic property of form but distinguished by individual formal idiosyncrasies which prevent their being collapsed with one another' (p. 104).

[2]The point is that speakers/hearers do not feel compelled to preserve generalizations in morphological processes. Fragmentation can be seen as replacing one generalization with several as the result of sound changes which obscure the original generalization. The fact that sound changes are not resisted when they threaten a generalization can be taken as evidence for one of two accounts of sound change. (1) The generalization is not really threatened because the lexical representations are abstract enough to preserve the general pattern, or (2) speakers/hearers do not need big generalizations, because little ones (particularized to sets of lexical items) will do just as well. This last view (which will be argued for in this paper) entails that lexical items are more important semiotically than are morphological processes.

[3]So Prokosch §56 'essentially a direct continuation of the IE perfect tense.' The singular endings in Gmc. are reconstructed as: *-a, -tha, -e* (compare Grk. perfect forms οἶδα, οἶσθα, οἶδε). The 2sg. in W. Gmc. probably comes from a different IE source and will be considered in the next section with the pret. pl. forms.

[4]In W. Gmc. the pret. 2sg. form also has the ablaut vowel which appears in the pret. pl. forms.

[5]'It is tempting to assume, though not definitely demonstrable, that Prim. Gmc. *ǣ > ē* in the form of West Gmc. from which OE and OFris. were derived ... and that this *ē* was then subject to change in two directions, becoming *ē* before nasal consonants, *ǣ* (or *ē*) elsewhere. Such a double development of *ē* would be parallel to the OE and OFris. treatment of *a*' (Campbell, 129). *ē > ō* is also more phonetically plausible than *ǣ > ō*.

[6]According to this theory the long vowel is the result of a conflation of two syllables which result from reduplicaton (which is a typical marker of perfect forms in IE). 'Thus from the root *sed*,sit, the IE perf. pl. would be **sesed- > *sēsd- > *sēzd-* > Gmc. **sǣt-*' (Campbell, 1959, p. 305, n. 1).

[7]'It expresses pure passivity, not necessarily passivity in past time' (Campbell, 727).

[8]The pass. part. suffix in Gmc. was *-an*. The three other suffixes which occur on stems with *u* (pret. sg. *-iz*, pret. pl. *-un*) and subj. pret. pl. *-in*) all had high vowels in Gmc. If the sound change (*u > o* before a low or mid vowel) occurred before these suffix vowels were lowered then the appropriate environment existed only for the pass. part.

[9]One has to wonder if any stage of any language can be found for which there is <u>not</u> evidence of fragmentation. Of course there can be local unity (ie. groups of lexical items which are treated similarly) but the trend seems always to be toward lexical specificity rather than morphological generality. From a diachronic perspective it seems to be fragmentation all the way down.

[10]*ie* in the present is the result of the influence of the palatal consonant *c*.

[11]*æ* in the present results from a failure of umlaut across the cluster *pp*. The pass. part. form may be due to analogy with the present.

[12]This dictionary is more concise than the OE dictionary that I used and so there may have been a bit higher rate of retention than I am reporting here. The concise dictionary was more convenient to use and gives a good general idea of retention rate.

[13]Moder(1986) seems to indicate that the vestiges of strong verb ablaut patterns are still salient and extendable.

[14]In this view of what a lexical representation is it would be surprising if there weren't lexical representations for all of the forms in a paradigm. This is because each separate form (as it is encountered in language use) is an event to be remembered. That the

memories of the forms in a paradigm are related to each other is a given. Exactly how this is done is a matter for future research. Also, note that this view of the lexicon is similar to Bybee's (1985, pp. 111 ff.) 'dynamic model'.

[15]The evidence presented in Esper (1973, chapter 6) indicates also that a pattern of relations among words can lead to the invention of new forms. He points out the necessity of treating both the analogical creation of new forms and the analogical revision of old ones as instances of the same type of process. As regards the conception of lexical representation which I am adopting here, I am sure that he would reject my 'subjective mentalism' and prefer to talk only about those aspects of language that are observable (ch. 7).

[16]This reinterpretation of o-vocalism as a marker of pass. part. is an instance of abductive change (Andersen, 1972).

[17]This order, if kept, would require a substantial revision of phonological theory.

[18]It could be salvaged as a synchronic rule by complicating it with an optional *r* between the vowel and the nasal. However, counterexamples of this revision can be found from *forma* 'first' to *terma* 'end'.

[19]This raises the whole issue of speech perception with abstract representations. The counter-argument to my claim here is that upon hearing Lat. *scrîbere* OE hearers 'heard' /screib-/ and thus identified the verb as class I. The problem with this is that with abstract lexical representations OE hearers had to sometimes 'hear' [ī] as /ei/ and sometimes as /ī/. Now, if they had a paradigm for *scrîbere* in which there was evidence for a stem like /scr_ib-/ it might be possible to claim that they 'heard' /ei/ in the present tense. In the absence of such evidence we must assume that they 'heard' /ī/ and thus that their lexical representations for class I verbs also had /ī/.

References

Andersen, H. 1972. Abductive and deductive change. Lang. 40:765-793.

Borden, A.R. 1982. A Comprehensive Old-English Dictionary. Washington: University Press of America.

Bybee, J.L. 1985. Morphology: A study of the relation between meaning and form. Amsterdam: J. Benjamins Publishing Co.

_____. 1986. On the nature of grammatical categories: A diachronic perspective. ESCOL'85. pp. 17-34.

Campbell, A. 1959. Old English Grammar. Oxford: Oxford University Press.

Esper, E.A. 1973. Analogy and association in linguistics and psychology. Athens: University of Georgia Press.

Janda, R.D. and B.D. Joseph. 1986. One rule or many? Sanskrit reduplication as fragmented affixation. ESCOL'85. pp. 103-119.

Kiparsky, P. 1968. Linguistic universals and linguistic change. in Bach, E. and Harms, R.T. (eds.) Universals in linguistic theory. New York: Holt, Rinehart & Winston.

Lass, R. and J.M. Anderson. 1975. Old English Phonology. Cambridge: Cambridge University Press.

Levin, S.R. 1964. A reclassification of the Old English strong verbs. Lang. 40:156-161.

Moder, C.L. 1986. Productivity and frequency in morphological classes. ESCOL'85. pp. 176-185.

Prokosch, E. 1939. A comparative Germanic grammar. Philadelphia: LSA.

Stratmann, F.H. & H. Bradley. 1891. A Middle English dictionary. Oxford: Oxford University Press.

OSU WPL 34 (1986) 123-126

The Etymology of bum: Mere Child's Play*

Mary E. Clark
The Clintonville Academy & Otterbein College
and
Brian D. Joseph

The etymology of the word bum in the meaning 'buttocks, bottom' is generally considered to be uncertain (e.g. by OED: s.v., by Onions et al. (1966: s.v.), and others).[1] One of the leading possibilities, though, is that bum is somehow a contraction of bottom (so Century Dictionary & Cyclopedia (1906: s.v.), Partridge (1966: s.v.), Morris and Morris (1977: s.v.)). This etymology has been denied, however, by the OED for two reasons: phonetic difficulties in the development from bottom to bum and the "historical fact that 'bottom' in this sense is found only from the 18th cent." (p. 1173), while bum, in its Middle English form bom, occurs as early as the 14th century (in Trevisa Higden Rolls, from 1387).[2]

There is little to say about the attestation question; we note only that attestation is often merely a matter of chance and it is quite possible that bottom referring to 'buttocks' might have been omitted from Middle English texts now available for reasons other than its nonoccurrence in the usage of time.[3] Also, the shift in meaning from 'bottom (in general)' to 'bottom part of a seated person' seems natural enough to have been able to have occurred independently at several times in the history of English.[4]

The phonetic "difficulties", however, can be dealt with, and it is that side of the etymological connection of bottom with bum that we wish to address here. It seems that the main objection to the phonetic derivation of bum from bottom stems from the fact that such a reduction or contraction was not a regularly occurring process in the historical phonology of English. In particular, bottom has continued into Modern English alongside bum, giving the current doublet, and there are words with a similar phonetic shape which have not undergone this "reduction", such as bottle and bodice, both attested early enough in English to be relevant to the matter at hand.[5] Thus if bum is derived in some way from bottom, it would have to have arisen in a dialect other than the one(s) providing the main input into standard Modern English.

A solution to these difficulties was suggested to us through observation of the usage of our older son David. At the age of 2 years 4 months (in late 1982), we heard David, while being diapered, say [ba(ʔ)ɔm], with a clear reference to the part of his anatomy we were most concerned with, i.e. his bottom. The glottal stop was somewhat weakly articulated, so that it was almost absent to our ears, and in fact later repetitions of the word may not have even contained it (hence the parentheses in our transcription). The resulting utterance sounded remarkably like adult bum. Moreover, the process responsible for the reduction evident in David's pronunciation of bottom seems to have been a regular one in his speech at that time.[6] During approximately the same period in his development, we heard [bawəl] for bottle, [lɪɪəl] for little, and [pɹmbeyɚr] for Paddington Bear (with assimilation of n to m), all with a medial dental stop[7] "reduced" and the resulting word "contracted", with some alteration of the vowels, when compared with the adult version. The regularity of this process in David's speech is shown also by the fact that at a later stage of development, 2 years 10 months, all of the above words which

had appeared in "reduced" form six months earlier came to have a medial dental stop in them.

Thus it is evident that there are "dialects"--or, more accurately, sociolects--in which the phonetic reduction of bottom to something like bum is perfectly regular. We propose, then, that adult English bum has its origins in child speech, especially in the relatively early stages of its acquisition. The occurrence of bum in adult English would then result from a form of dialect borrowing, fostered by the opportunity for close and frequent adult-child interaction that diapering provides. The parent would thus be using and incorporating into his/her own speech a true child language form.

As with any borrowing--dialect or otherwise--or neologism, the entry of bum from child language into adult speech needs only to have occurred once, though the possibility of recurring borrowing at several points in the history of English cannot be discounted.[8] Once a part of the mainstream dialect, the retention and spread of this word becomes a matter not of dialect borrowing but instead of the regular lineal transmission of language through subsequent generations. However, one interesting aspect to this proposed borrowing--whether it occurred once or many times--is that it is entirely in keeping with the Neogrammarian view of sound change, in which dialect borrowing can be an explanation for apparently irregular sound changes in a given speech community. Here the donor dialect--child language--had the phonetic reduction regularly and the borrowing into adult language led to the bum/bottom doublet and the seemingly irregular and sporadic sound change linking the two.

Moreover, parallels can be found for the type of development suggested here for bum. The word tummy, for instance, is universally accepted (e.g. by OED, the American Heritage Dictionary, Partridge, etc.) as being in origin a nursery form or infantile alteration of stomach; its use by adults is especially common when they are talking to children, but it has penetrated into adult speech sufficiently to form the basis for a product name (Tums) and advertising slogan (Tums for the tummy). Similarly, bye-bye, as observed by Dilkes 1983, is another nursery word--here probably an adult conventionalized child form rather than a form originating with children--which has made its way into adult use.

Notes

[8]We would like to thank Zheng-Sheng Zhang of the Department of Linguistics of the Ohio State University for his help with some of the research on this paper.

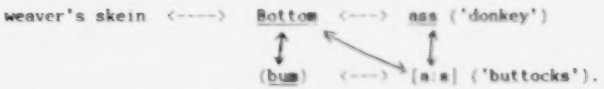
1. Some sources venture no opinion at all; Webster's 3rd, for example, has no comment on the etymology of this word.

2. The actual citation with the one attestation runs thus:

> He hadde many ȝere ƿe evel ƿat hatte ficus, ƿat is a schrewed evel, for it semeƿ ƿat his bom is oute ƿat haƿ ƿat evel. [6.357]

This is apparently the only Middle English occurrence of bom, for both the MED and the OED list only this lone example.

3. It is possible that the OED was somewhat precipitate in its pronouncement that bottom meaning 'buttocks' dates only from the 18th century. We suspect that in A Midsummer Night's Dream (1595), the name of character Bottom may be part of an elaborate Shakespearean pun which plays upon a meaning of 'buttocks' for common noun bottom. In the first place, Bottom is given the head of an ass (donkey); with the contemporary (American) senses of Bottom and Ass, i.e. bottom = 'buttocks' and ass ='buttocks; donkey', there is a reasonably good pun playing on Bottom's name and his fate. The double sense of ass, though, is the result of a sound change merging ass 'donkey' with the ME and ENE arse 'buttocks' through the loss of preconsonental r (with some vowel changes as well, most likely). However, there is a strong possibility that arse had an r-less pronunciation as early as Shakespeare's time. Barber (1976: 319) points out that "/r/ was lost in some non-standard forms of speech in lME, especially in eastern dialects and in substandard London speech ... before /s/ and /ʃ/". For example, Barber continues, "in Troilus and Cressia, Shakespeare uses the word tercell 'a male hawk' [while] in Romeo and Juliet, the same word occurs in the form tassell". Both r-ful and r-less forms of arse are likely to have coexisted in Shakespeare's time, increasing the possibility of a play on arse/ass. Now, according to Rowse (1978: I.232), Bottom's name is already to be understood as a pun on his occupation as a weaver, bottom referring to "a skein on which the weaver's thread is wound" (and we note also that weavers of necessity spend a lot of time sitting); it therefore is not unreasonable to suggest that Shakespeare, as an ardent and often ribald punner, may have intended a double pun, pivoting on the relationship between Bottom's name and his identity with ass (= 'donkey' and 'buttocks'). Diagrammatically, then, the relations in these puns are:



The success of these puns depends on a meaning of 'buttocks' for both bottom and ass (through the near-homonym in the nonstandard pronunciation of arse). Thus we conclude that the meaning 'buttocks' for bottom was available somewhat earlier than the OED suggests, at least as early as the late 16th century.

4. David Stampe of the Department of Linguistics of University of Hawaii has informed us that he has found virtually the same semantic connection in the Munda languages of India that he has worked with, adding plausibility to our claim that it is indeed a natural connection. We note also that among the meanings given for Middle English botme in the MED are several from which a shift to the meaning 'buttocks' would be fairly straightforward; especially relevant are the meanings 'the part of a bodily organ farthest from the exterior; bottom of the stomach' (p. 1077).

5. The putative "reduction" involves elimination of the medial consonant as well as alteration of the vowels. The source we propose below addresses both of these matters.

6. We know of several other mothers who have also noted a similar pronunciation of bottom from their children, making it likely that such a reduction is characteristic of children's speech in general.

7. We hesitate to label this stop, for it is unclear to us whether the input to David's speech included a /t/, /d/, or flap /D/ in these words.

8. This is especially true if the reduced form is a common child language pronunciation; see footnote 6.

References

American Heritage Dictionary of the English Language. (1978). Wm. Morris, ed., Boston: Houghton-Mifflin Co.

Barber, C. (1976). Early Modern English. London: Andre Deutsch.

The Century Dictionary and Cyclopedia. (1906). New York: The Century Co.

Dilkes, M. (1983). "Bye-bye", American Speech 58.92.

MED = Middle English Dictionary. (1956). H. Kurath & S. Kuhn, eds. Ann Arbor: University of Michigan Press.

Morris, W. & M. Morris. (1977). Dictionary of Word and Phrase Origins. New York: Harper and Row.

OED = The Oxford English Dictionary. (1961). J. Murray, et al., eds., Oxford, at the Clarendon Press.

Onions, C. T., G. W. S. Friedrichsen, & R. W. Burchfield, eds. (1966). The Oxford Dictionary of English Etymology. Oxford at the Clarendon Press.

Partridge, E. (1966). Origins: A Short Etymological Dictionary of Modern English. New York: MacMillan.

Rowse, A. L. (1978). The Annotated Shakespeare. New York: Clarkson N. Potter, Inc., Publishers.

Webster's 3rd = Webster's Third New International Dictionary of the English Language Unabridged. (1966). Springfield: G. & C. Merriam Co.

OSU WPL 34 (1986) 127-132

Small Group Lexical Innovation: Some Examples

Christopher Kupec

The slang of small groups is usually ignored or dismissed by writers of popular etymology books and slang dictionaries. Scientific linguistic journals as well do not often deal with this subject.[1] Yet it is a valid part of the linguistic competence of the individuals using this slang, and as such deserves some attention by linguists. For one thing, individuals, and not speech communities, are the creators of new words. And individuals create them within the setting of their own social groups. Moreover, if we study this locale of lexical innovation in more detail, we can discover a number of unusual processes of word-formation. We can also gain some insights into the social side of language change, and in particular, into the social mechanisms involved in the creation, acceptance, and usage of innovative words.

Some examples of small group lexical innovation are provided in this paper in order to undermine the notion that slang uses only typical processes of word-formation (e.g., compounding, prefixation, suffixation, semantic radiation, concatenation, etc.), and to investigate the social side of word-making.

The following examples were coined and are used by a circle of ten or so close friends. All members of this group are in their early twenties, although some of the words they've created originated during their high-school years. Where possible, the various creators are identified (by initials), in order to make it easier to see who creates most of the words.

1. Orthographic Origin

bicth [bIkɛ̰] n. 'a person' Used jocularly and derogatively. Created when M. attempted to write bitch backwards on the inside of a steamed-up car window, so it could be read from the outside. Note that the change affected the meaning melioratively. The [kɛ̰] cluster is nonexistent in standard English (Cohen 1962: p.74).

said-ass case [sɛd æs keis] n. 'a confused, pathetic individual', from a spelling error in a letter that M. wrote to another member of the group.

music for monkey n. 'music' A spelling error of a songtitle, 'music for money', written by M. on a cassette case.

nelk neek [nɛlk nik] interj. A nonsense, identification word.[2] Originally neek-neek, an interjection used by Jack Nicholson's character in Easy Rider after taking a swig of whisky. M. had the word ironed onto a T-shirt in captal letters, NEEK NEEK, and after several washings the top two horizontal bars of second E wore off, leaving NELK NEEK. Created by V.

nice teds n. 'nice breasts' A misspelling of nice tits by a Palestinian friend of the group, written on a Valentine's Day card for little children. It's possibily his phonemic representation of a word he'd never seen in print.

2. Phonetic Fortition

jooj; joojeh [džudž]; [džu·džʌ] n. 'a person; a guy' Sometimes used as a nonsense, identification word. D.'s emphatic pronunciation of dude, an unpopular word in this group. Jooj or its variant form joojeh, however, is acceptable. Possibly this is because it isn't dated or common, and because it disparages the people who do actually say dude. The phonetic changes appear to be fortitive, since [d] is turned into an affricate [dž], which is very expressive (as evidenced in Jesus, jeez, jerk, and jism), and extremely forceful. The variant joojeh seems to have arisen from a reinterpretation of the vocalization after the release of the second affricate as a full vowel. Note that joojeh has its stress on the second syllable, underlining the fact that it is a fortition.

loozar [lu·zar] n. 'an imbecile; a moron' M.'s pronunciation of loser. The r-colored schwa in loser has been reinterpreted as [ar] when both syllables are given equal stress. From this, it would seem likely that /ar/ is M.'s phonemic representation of the agentive -er ending. A variant exists in a front-clipped form, zar.[3]

hroach; heroach [hrouc]; [h rouc] interj. 'I got burned' or 'You got burned (by the speaker or someone else)' M. and G.'s pronunciations of roach, 'burnt' or 'ruined'. The origin of the initial [h] is uncertain.[4] In the variant heroach, the transition from the word-initial [h] to [r] has been reanalyzed as a vowel, possibly because of the trouble in pronouncing [hr] as a word-initial consonant cluster in standard English.[5]

skrch [skrč] interj. 'scratch; ouch' Created by V. Said by him to sound more onomatopoetic than scratch.[6]

3. Front Clipping

splud [splʌd] n. 'a kidney punch' & interj. 'take that!' Created by M. by clipping 'Piss blood!' It's apparently unrelated to the British English curse, 'sblood for 'Christ's blood'. Notice how the removal of the verbal element brought on the treatment of splud as a noun.

medic [mE·dIk] interrogative. 'want my dick?' Created by K. for use in public by clipping 'Want my dick?'. Along with the pronunciation, the stress changed as well to coincide with medic meaning 'corpsman'.

zar [zar] n. 'a jerk' See loozar above.

4. External Source

hojoba [hou·dž›·bə] interj. An identification word. It comes from jojoba, an ingredient in certain shampoos. The pronunciation however is neither straight from the Spanish word jojoba nor a direct Anglicization ([ho·ho·ba] and [džow·dža·bə], respectively). Rather it is a merger of the two. Originated with K. and N..

neek neek [nik nik] interj. An identification word. Originated with M. See nelk neek above.

5. Word Manufacture Ex Nihilo

weep wa, weep wawl [wip·wa]; [wip wa̰l] interj. An identifacation word. Believed to have been created by N.'s brother-in-law.

woob [wub] vt. 'to deceive, toy with', vi. 'to wobble, waver', & interj. M. originated it, but he doesn't remember how. The verb is completely regular. The intransitive verb's meaning seems to have been influenced by the similarly spelled word wobble. WOOB was what M. had ironed onto the other side of his NEEK NEEK T-shirt.

schweibik [ʃwei·bik] interj. An identification word. Created by N. (along) with schwebel, scheben, and schlabonowitz) while daydreaming in his German class. He wanted German-sounding nonsense words, although he wasn't sure if they meant anything in German. But it seems that they're nonsense words in German as well.[7] The four words are often used in a catch-all phrase, 'Schwebel, schweibik, schleben, schlabonowitz', which is spoken as if it were a verb conjugation.

schwebel [ʃwei·bl̰] interj. An identification word. See schweibik above.

schleben [ʃlei·bn̰] interj. An identification word. See schweibik above.

schlabonowitz [ʃlɔ·ˀbə·nɔ·wɪc] interj. An identification word. See schweibik above.

6. Malapropism

dispose v. 'to suppose; to guess' A play on suppose, created by M.

reconcile v. 'to guess, to reckon' Created by M. from reckon.[8]

7. Semantic shift

Jack jizz pn. A nickname for M.'s father. Originally Jack jizz meant M. (the reasoning being that in common slang, jizz means 'semen', so Jack jizz meant 'Jack semen', i.e., M. whose father's name is Jack). But through time, the meaning shifted back to 'Jack', so the second half of the nickname is just added baggage.[9] The reasons why this happened are unclear. Jizz was tried as a regular patronymic in this group, but only Jack jizz has survived, however, so perhaps it is because of its alliteration, and its being influenced by expressions like Jackshit and Jackdick.

Nelson pn. A nickname for K. Since another of K.'s nicknames is The legend, this nickname is a hidden pun. It is dependent on one's knowing the name of a park in northeastern Ohio called "Nelson's Ledges" or 'Nelson Ledges', which sounds similar to Nelson Legend.[10]

Wisner Road Resident n. 'a melonhead' Specifically it is a nickname for B. Where this group comes from, there is a local legend about melonheads: inbred, insane albinos with huge heads who come out at night to wreak havoc on anyone foolish enough to drive down Wisner Rd., where supposedly these creatures live. Calling someone a Wisner Road Resident is the same as calling him a melonhead.

The Nightmare pn. A nickname for M.'s old Plymouth. When Lee Iacocca started going on TV to advertise his company's cars, there was a joked-about phrase, 'a Lee Iacocca dream-car', for any car made by Chrysler. M. owned an old Plymouth Fury that was a piece of junk, so instead of calling it 'a Lee Iacocca dream-car', he called it a Lee Iacocca nightmare, or for short, The Nightmare

B. Metathesis

kirsheds [krʌ́ɛc] n. 'crushed red pepper' Created by K. Actually this was a speech error, but K. insists that he meant to say it, and this could be true, since crusheds has never been used by this group. If crusheds were pronounced [krʌ́ɛc], it would seem easier to say kirsheds [krʌ́ɛc], since [kr sc] is a more complicated articulatory sequence than [krʌ́ɛc] (i.e., back to front to back to front vs. back to front).

The processes of word-formation illustrated here emphasize the variety of methods available and the prevalence of unusual paths in lexical innovation. Orthographic origin, fortition, and malapropism are given little or no status in most published works on word-formation. Sentential front-clipping is not a common process either. As for metathesis, semantic shift, external origin, and word manufacture ex nihilo: how often do we get to see how a word comes about, why it comes about, and who is responsible for it?

The examples shown also let us look in on the social aspect of lexical innovation. The prevalence of identification words (words that have no defined meaning but are used to strengthen group identity and cohesiveness) points out how important the social setting that these are created in is to the members of the group. Using such words, in a sense, affirms (both for the speaker and for the rest of the group) that the speaker is part of the group. This is the same reason that much technical jargon is used in various circles. Everybody wants to belong.

The humor of these words is also worth mentioning. New words seem to be accepted by the group based on their humorous content, as well on their productivity. If a word's not funny, it's not used. And once it loses its freshness, the word is dropped from use more often than not. This is because word-formation and usage are active, self-conscious processes. Quite often members of the group will comment on their irritation over a word: either that they use it too much or that somebody else does. This habit extends beyond the group. Parties and get-togethers are excellent locations to attack dated slang and "sloppy speech" used by strangers. If a word isn't liked, the attacker uses it on the person continually until he or she realizes what the problem word is. Sometimes this method will backfire and the word is instead considered to be well received. A number of words have been created that way in this group: kirsheds, skrch, and loozer for example. The group is constantly on the lookout for verbal deviance in themselves and in others outside of the group.

That all these mechanisms and methods need to be examined in greater detail is obvious. I'm not the first to notice a lack of literature and research on these topics. But, hopefully, this paper gives others some ideas and concrete examples concerning lexical innovation and its social setting.

Notes

1. Maledicta is a major exception to this trend.

2. Identification words are usually nonsense words that are used to fortify relationships within a group. They delinate who's who. Nonsense baby-talk used between lovers can also be included in this term, since it strengthens the bond between them and it affirms the loving nature of the relationship. Essentially, identification words identify the group that the speaker belongs to.

3. Zar works well in public, because the unknowing recipient of the insult can think he's being admired, cf. czar.

4. The #h- might have arisen from the heavy aspiration given to the [r]. The tongue is extremely retroflexed and often the pharyngeal fricative [ħ] is used or the uvular fricative [χ] instead of [h].

5. It might be worth noting that all the aspirated-r words that English took from Greek are unaspirated in English. But since the [h] of hroach was desired, a vowel insertion was made, so that the [h] wouldn't be lost.

6. As V. put it, 'Bees go "bzzz" not "buzz"' (from a conversation).

7. It is possible, however, that schweibik could be Schwäbig 'Swabian' (Swabia being a region in southwestern West Germany) in German. Schwebel is a surname, as in Schwebel's, which is a bread-baking company in the U.S. Schleben could have been influenced by secondary reduplication of the type, actor schmactor, i.e., leben schleben (since in German, leben means 'to live'). Schlabonowitz could be a surname also. All these words are probably influenced by the expressive, derogatory nature of sC- in Yiddish slang (e.g., shtup, schlang, and schlock).

8. Usually only used in the sentence, 'I reconcile...'. The same usage is used with dispose.

9. This type of semantic change is not infrequent. For example, rhubarb is derived from the Late Latin rha barbarum 'foreign rhubarb'. But when only one species was present, the 'foreign rhubarb', the classifying modifier lost its meaning. So maybe in the case of Jack jizz, the high number of nicknames for M. along with the lack of nicknames for his father influenced the change in meaning.

10. An example of the impression that the inexact pun is funnier than an exact one.

References

Bauer, Laurie. (1983). English Word-Formation. Cambridge University Press. Chapter 7.

Cohen, A. (1952). The Phonemes of English. The Hague: Martinus Nijhoff. p.74.

Jeffers, Robert J. and Ilse Lehiste. (1984). Principles and Methods for Historical Linguistics. Cambridge: MIT Press. Chapter 8.

Lehmann, Winfred. (1973). Historical Linguistics: An Introduction. New York: Holt, Rinehart, and Winston, Inc. pp.205-234.

Marchand, Hans. (1966). The Categories and Types of Present-Day English Word-Formation. University of Alabama Press.

Sturtevant, E.H. (1942). Linguistic Change. New York: G.E. Stechert & Co.

Wentworth, Harold and Stuart Berg Flexner. (1960). "Introduction to the Appendix", in Dictionary of American Slang. New York: Thomas Y. Crowell Co., pp.596-608.

OSU WPL 34 (1986) 133-141

WORD FREQUENCY AND DIALECT BORROWING

Debra A. Stollenwerk

1. Introduction

That word frequency plays a significant role in the spread of language change was suggested as early as a century ago by Schuchardt. (Phillips 1984: 321) In this century George Zipf (1929) has proposed relative frequency as a determinant of sound (phonetic) change. This frequency hypothesis claims that phonetic change (i.e. physiologically motivated change such as assimilations, vowel-reductions and the like) operates on high-frequency items first; all other change (characterized as analogical or conceptual--i.e. not physiologically motivated) affects low-frequency forms first (Phillips 1984:336-337).

That frequency is a significant factor in the spread of language change is supported by evidence from numerous studies. Leslau (1969) presents evidence in Ethiopian languages that certain phonetic changes such as spirantization, elision and assimilation affect high-frequency words first. Fidelholtz (1975) gives evidence that vowel-reduction (a phonetic change) in initial syllables correlates positively with frequency. For example, astronomy, mistake and mosquito, classified as relatively frequent forms, may occur with a reduced vowel in the first syllable; less frequent words such as gastronomy, mistook and Muskegon are less likely to occur with the reduced vowel--although it is noted that residents of Muskegon are more likely to reduce the vowel in that form presumably because it is a more frequent form for them. One might expect similar behavior for the item Sandusky (Fidelholtz 1975:200).

Hooper's (1976) analysis of schwa deletion in English (yet another phonetic change) yielded an identical conclusion. Consider the following word pairs: nursery/cursory; scenery/chicanery; celery/artillery; memory/armory. In each pair it is the more frequent form (i.e. the first form) which is more likely to undergo schwa-deletion.

The evidence also suggests that it is the low-frequency forms which are the first to undergo non-phonetically motivated change (i.e. conceptually motivated or analogical change). It is this type of change which is the subject of Toon's (1978) analysis of h-deletion in Old English morpheme initial consonant clusters written as hn, hr, hl and hw. Here, he found that the low-frequency words exhibited the greatest rate of deletion. In his discussion of lexical diffusion in early Old English, Toon presents a model of sound change in which an innovation operates initially on low-frequency items, gradually spreads to and accelerates through high-frequency forms to near-completion and in the final stage leaves a small residue behind after reaching completion--e.g. the merger of /eo/ and /io/ and the raising of West Germanic *a to /o/ before nasals. That sound changes may operate initially on low-frequency words suggests an explanation for "either the initial state of low-level variable application or a completed end state with residue" both of which are regularly observed in sound change. (Toon 1978:362)

Additional evidence of analogical change affecting low-frequency forms first comes from Hooper's study of the six verbs creep, leap, weep, leave, sleep and keep (Hooper 1976). Each of these verbs has a preterite form occurring with a lowered, laxed vowel--e.g. crept, kept, leapt. Only the first three, however, can optionally take a regularized form in the preterite--i.e. creeped, leaped and weeped. Analysis of frequency reveals that the mean for the forms subject to leveling is 37 while the mean for those not subject to leveling is 485. Thus, the leveling (analogical change) operates on the less frequent words.

While there seems to be a good deal of evidence in support of a frequency-effect on the spread of sound change there may also be evidence that frequency is a factor in dialect borrowing. What I propose to do in this paper is to examine the possible role of word frequency in dialect borrowing by presenting some evidence from my own dialect in which words that are, for the most part, derived historically from Middle or Early Modern English ō vary in pronunciation.

2. Variation in 'og'-words

There occur in my speech different phonetic realizations of graphic ō before /g/ such that graphic <o> is realized as either /ɑ/ or/ /. (It should be noted here that /ɑ/ and /ɔ/ are contrastive in my speech, serving to distinguish, for example, cot and caught or tot and taught.) Based on my auditory perceptions the distribution of these sounds before /g/ is as follows:

/ɔ/	/ɑ/
frog	cog
log	clog
fog	jog
hog	*smog
dog	bog
*smog	

Within this corpus of data all forms except smog and jog are derived from either Middle English or Early Modern English ō--thus,

frog < ME frogge
log < ME logge
fog < ME fogge
hog < ME hogge
dog < ME dogge
clog < ME clogge
cog < ME cogge
bog < OE bugan, Early Modern bogge

(from the Oxford English Dictionary, 1933). Because of the common phonetic source of the vowels of these words (i.e. Early Modern or ME ō) it might be expected that the graphic <o> would be homophonous. The items fog and smog might be expected to be homophonous as well since smog, a relatively recent word, (the first attestation being 1905) is derived from fog (via a blending with smoke). The word jog appears to be relevant to the data base as well although its etymology is uncertain (possibly an alteration of ME shog or derived from Early Modern iogge (Oxford English Dictionary 1933).[1]

smog is found with variable pronunciation, perhaps a reflection of the confusion I am currently experiencing over the pronunciation of this word. (In fact, in a recent lecture, I said the word twice. The vowel of the first utterance I perceived as rounded while that of the second as unrounded. Moreover, the second utterance was perceived by me as a correction of the first.)

Spectrographic analysis of the syllable nuclei in these forms supported the intuitive distribution (see above) with the exception of smog which exhibited some rounding. The words, on the basis of spectrographic evidence, have been arranged into three groups: (A) consists of words in which the phonetic realization was /ɔ/; (B) consists of forms in which the phonetic realization was /a/; (C) consists of smog in which the intended target (i.e. the target at which I perceived myself to be aiming) was /a/ but which spectrographically exhibits some rounding. The values for F1 and F2 for each item are listed, as well as the mean (X) and standard deviation (SD) for each group.

Table 1

	Item	F1	X / SD	F2	X / SD
(A)	frog	769.3		1153.8	
/ɔ/	log	846.2	X=800	1153.8	X=1199.9
	fog	769.3	SD=42.1	1153.8	SD= 68.78
	hog	846.2		1230.7	
	dog	769.3		1307.6	
(B)	cog	923.1		1653.9	
/a/	bog	1000.0	X= 897.46	1615.2	X=1653.83
	jog	769.3	SD= 95.91	1692.4	SD= 31.52
(C)	smog	846.2	X=846.2	1461.4	X=1461.4

Lip-rounding and labiality have the effect of lowering formants. Clearly, then, the items in (A) exhibit a more rounded vowel than those in (B) as evidenced by the higher F2 values of the forms in (A). In smog, where the intended target (intuitively speaking) was /a/, the vowel is immediately preceded by a labial consonant. F2 for these items is noticeably lower than the values for F2 in (B).

So, group (A) exhibits roundedness with a mean of 1199.9 for F2; (B) clearly exhibits less rounding with a mean F2 of 1638.47 and (C), where intent and realization diverge, is intermediate between (A) and (B) with a mean F2 of 1358.86.

In (B), the item bog also contains a labial consonant immediately preceding the vowel, yet there is no significant lowering of formants. Because of its stop-quality, however, /b/ is more unlike a vowel while /m/, being a sonorant, is more vowel-like. Given the intended target of /a/ for (C) as well as the intermediate F2 value, it would appear then that co-articulatory rounding is a factor in the realization of the vowel in (C).

In order to analyze overlapping I plotted on a graph the X-value for each group as well as two standard deviations above and below that point. The resulting graph showed that the phonetic realization of smog more closely approximates the rounded vowel than the unrounded vowel, probably a result of co-articulatory rounding.

3. Frequency effect

Referring again to the word groups of Table 1 (shown below for convenience) and temporarily ignoring group C, a pattern may be observed in terms of the frequencies of items within groups A and B.

(A)	(B)
frog	cog
log	bog
fog	jog
hog	clog
dog	

That is, the data suggest that the items of (A)--realized with the rounded vowel--are of higher frequency than those of group (B)--realized with the unrounded vowel. (The item in (C) is intermediate between (A) and (B) in terms of roundedness and appears to be undergoing co-articulatory rounding as a result of a preceding labial sonorant.)

In order to examine the issue of frequency, two sources giving a frequency analysis of English were consulted. The Kucera and Francis volume (1967) ranks items by means of a three-figure number--e.g. 1-01-001--where the first figure designates frequency of occurrence in the corpus; the second figure indicates the number of genre subdivisions (out of a total of 15 genre subdivisions) in which the word occurs; and the third figure indicates the number of samples (out of a total of 500 samples) in which the item is found. What follows then is a ranked listing of the data in order of least frequency to greatest frequency:

bog	1-01-001
cog	1-01-001
frog	1-01-001
jog	1-01-001
smog	1-01-001
clog	2-02-002
hog	3-02-003
log	11-05-007
fog	25-09-018
dog	75-12-028

The listing shows a clear delineation of low-frequency/high-frequency between hog and log and indicates as well that hog and frog (counter to my intuition) are of relatively low frequency. These frequency counts, however, are based on written usage rather than spoken usage. And, as Hooper (1976: 98) notes, frequent forms in written text are found to occur even more frequently in spoken usage while less frequent forms in written text occur even less frequently in spoken usage.

The American Heritage Word Frequency Book ranked 86,741 different words out of a total corpus of 5,088,721 tokens. What follows, again, is a ranked listing from least to greatest frequency where F represents total occurrence of the form in the overall corpus and U represents the estimated frequency per million tokens:

	F	U
groggy	2	.1142
cog	2	.1152
clog	2	.1257
jog	3	.2477
bog	14	1.7289
smog	22	2.4598
hog	33	5.0096
frog	171	26.143
fog	212	33.553
dog	1380	231.49

The U-figure of the righthand column shows a fairly sharp increase between smog and hog (slightly more than a 100% jump). If the line between low and high frequency is drawn here, there is accord between intuitive judgments and interpretation of frequency data. In any event, the item frog is of notably higher frequency in this analysis than in that of Kucera and Francis.

Analyzing, then, the distribution of /a/ and /ɔ/ in these forms in terms of frequency (the lexical item hog being the only relatively infrequent form in group (A)), there may be reason to assume that the variation in pronunciation among the relevant forms is linked to a frequency effect; that is, the frequent forms (with hog being interpreted as fairly frequent) exhibit the rounded vowel while the relatively infrequent forms exhibit the unrounded vowel.

4. Dialect borrowing

The frequency effect which I am proposing here differs somewhat from the frequency effect of the aforementioned studies by Hooper, Phillips, etc. That is, the focus of those particular studies was the role of frequency in language change (be it phonetic or conceptual change) whereas the focus of this paper is the role of frequency in dialect borrowing. It is possible that the variation under analysis in this study represents language change in its initial stages. However, this type of change seems to be neither phonetically motivated nor analogical. Possibly this variation does not so much represent sound change motivated by internal factors as dialect borrowing motivated by external influence. The latter is clearly a possibility given the influence which various dialect areas have had on my speech--namely, Toledo, Ohio (birth to age seven); Jacksonville, Florida (age seven to ten years); Portland, Michigan (age ten to sixteen years); and Columbus, Ohio (age sixteen to the present).

Marckwardt, in his study of Middle English ō in the Great Lakes region (1940:570), identifies the unrounded vowel as primarily a Northern feature (i.e. in my case, a feature of Michigan speech) and reports the following distribution for the stressed vowel in fog, foggy, frog and hog: /a/ prevails

throughout Michigan and /ɔ/ is found throughout Ohio. For the items log and dog /ɔ/ is favored throughout the entire Great Lakes region. (1940: 562) Clearly, then, in view of the above distribution, the /ɔ/ quality of the stressed vowel in fog, foggy, frog and hog represent an Ohio (or non-Northern) pronunciation.

While the distribution of /ɔ/ and /a/ in Michigan and Ohio is divergent for some forms, the status of ME ǫ́ in many groups of lexical items in the Toledo area is identical to that of Michigan. For example, in the items hospital, vomit, college, cottage, stopper, pocket, God and hod the stressed vowel in each is unrounded in both northern Ohio and Michigan while being rounded in Central Ohio (Marckwardt 1940:566-569). However, this area of Northwestern Ohio has been analyzed as a transition area by Davis and McDavid--that is, an area which has experienced (or is experiencing) influence from two or more "directions so that competing forms exist in it side by side (1960:264). They found, in their five-county survey, that the unrounded vowel was centered largely at Perrysburg (a suburb of Toledo). However, even here variable pronunciation of the vowel in fog, foggy, on, pa and grandpa occurred (19:270).

Distribution of these vowels was not available for clog(ged), smog, bog(ged), jog and cog. Therefore, in order to determine the pronunciation of the syllabics in these forms for the Toledo area (which in my speech occur with the unrounded variant) I selected two natives of Lucas county who read a set of sentences containing these forms (each sentence being read twice) from which spectrograms were made. (One informant was, in fact, a native of Perrysburg.) Spectrographic analysis yielded the following distribution:

Table 2

Distribution of /a/ and /ɔ/ in Toledo

	Inf. 1		Inf. 2	
smog	/ɔ/	/ɔ/	/ɔ/	/ɔ/
bog(ged)	/ɔ/	/ɔ/	/ɔ/	/ɔ/
jog	/a/	/a/	/a/	/a/
cog	/ɔ/	/ɔ/	/a/	/a/
clogged	/ɔ/	/ɔ/	/ɔ/	/ɔ/

Note that the only item which was consistently unrounded in the speech of both informants is jog and that one informant pronounced cog with the rounded vowel while the other produced this form with the unrounded vowel. These results are consistent with the findings of Davis and McDavid in terms of the existence of competing forms in the area.

Thus the status of graphic <o> before /g/ based on my own findings as well as those of Marckwardt, Davis and McDavid can be condensed into the following table:[2]

Table 3

		Michigan	Toledo	Self
	log	ɔ	ɔ	ɔ
	dog	ɔ	ɔ	ɔ
1)	fog	a	a/ɔ	ɔ
	foggy	a	a/ɔ	ɔ
	frog	a	ɔ	ɔ
2)	hog	a	ɔ	ɔ
	smog	a	ɔ	a/ɔ
	bog(ged)	a	ɔ	a
3)	jog	a	a	a
	cog	a	a/ɔ	a
	clog(ged)	a	ɔ	a

The table is divided into three groups of items--i.e. 1) high-frequency; 2) intermediate-frequency; and 3) low-frequency. The results indicate that /a/ prevails in Michigan regardless of frequency and predominates in my speech in only the low-frequency forms whereas / / prevails in the speech of the Toledo informants. Note also the clear divergence of my speech from the Toledo pattern in the words bog and clog where my speech exhibits the unrounded (Northern) vowel and that of the Toledo informants invariably exhibits the rounded vowel. What appears to be happening here is outside dialectal influence--or dialect borrowing.

Labov (1972) discusses the dialect borrowing/restructuring which may occur when a speaker with an already-formed linguistic pattern moves into another dialect area. He identifies the formative period of first language acquisition as four to thirteen years of age and in his study of New York City's vowel system in the speech of informants from the Lower East Side found ten years to be the critical, cut-off age for native speakers moving into New York from other U.S. dialect areas (1972:305). Thus, one would expect that a speaker who moves into New York before age ten is more likely to adopt the vowel pattern of New York than a speaker who moves into New York after age ten, the age at which I moved to Michigan.

Given the the evidence from Labov's studies and the predominance of the unrounded vowel (a Northern feature) in only the low-frequency forms in my speech, it seems reasonable to propose that sometime between age ten and sixteen I adopted the vowel pattern of Central Michigan. Clearly, however, this did not represent a wholesale adoption since higher-frequency forms such as fog and hog, which in Michigan exhibit an unrounded vowel, in my speech retain the rounded (Ohio or non-Northern) vowel. Thus, it would seem that word frequency is playing a significant role here in the phenomenon of dialect borrowing.

Antilla (1972) makes the claim that not only does frequency play a role in language change but in pronunciation borrowing as well. "Words with high local frequency tend to be the last ones to be changed...[because]...high local frequency acts as a barrier to change from the outside." (1972: 188) In this case, his claim would mean that high frequency words such as fog, log and dog, whose forms are firmly established in the memory of a speaker, would be resistant to influence from outside dialect patterns. Low frequency words, on the other hand, such as clog or bog, whose forms are not as firmly established in the speaker's memory, are less stable and therefore more likely to be affected by outside change.

5. Conclusion

It is this, then, which I propose here--that the phonetic variation found in what might reasonably be expected to be a homophonous corpus (with respect to graphic <o>) can be explained in terms of the role which word frequency plays in the process of dialect borrowing. That is, the high frequency forms (with the unrounded vowel) appear to have resisted the influence of a Michigan vowel pattern whereas the less frequent forms (with the possible exception of hog which may actually be of higher frequency in spoken usage) were more susceptible to outside influence. It may well be that word frequency figures significantly not only in language change but in synchronic variation as well.

Notes

*I am very grateful to Keith Johnson for the hours of patient listening and assistance he gave me--especially with word-processing and phonetics dilemmas. Hopefully, the incessant interruptions are over.

1. The corpus might reasonably be expanded to include other 'og'-words such as soggy, groggy and eggnog which, even though not historically derived from ME o, are graphically identical and may be rhymed with, for example, foggy.

2. The results for items smog-clog(ged) in (Central) Michigan are based on my own auditory judgements and are consistent with the predominance of /a/ in that area for not only the forms in 1) and 2) but graphic-<a> words like want, watch, ma and pa and other graphic-<o> words such as vomit, on, cottage and pocket as well (Marckwardt 1940).

References

Carroll, John B. 1971 (ed.) American Heritage Word Frequency Book. New York: Houghton Mifflin.

Davis, Alva L. and McDavid, Raven I. 1950. Northwestern Ohio: A transition area. Language 26.264-273.

Fidelholtz, James L. 1975. Word frequency and vowel reduction in English. Chicago Linguistic Society 11.200-213.

Hockett, Charles F. 1958. A Course in Modern Linguistics. New York: The MacMillan Company.

Hooper, Joan B. 1976. Word frequency in lexical diffusion and the source of morphophonological change. Current Progress in Historical Linguistics 96-105

Hudson, R. A. 1980. Sociolinguistics. Cambridge: Cambridge University Press.

Jeffers, Robert J. and Lehiste, Ilse. 1979. Principles and Methods for Historical Linguistics. Cambridge, Massachusetts: MIT Press.

Kucera, Henry and Francis, W. Nelson. 1967. Computational Analysis of Present-Day American English. Providence, R.I.: Brown University Press.

Labov, William; Malcah Yaeger; and Richard Steiner. 1972. A Quantitative Study of Sound Change in Progress. Philadelphia: The U.S. Regional Survey.

Labov, William. 1981. Resolving the Neogrammarian controversy. Language 57.267-308.

_____.1963. The social motivation of a sound change. Word 19.273-310.

_____1972. Sociolinguistic Patterns. Philadelphia: University of Pennsylvania Press.

Leslau, Wolf. 1969. Frequency as determinant of linguistic changes in the Ethiopian languages. Word 25.180-89.

Marckwardt, Albert H. 1940. Middle English o in American English of the Great Lakes area. Papers of the Michigan Academy of Science, Arts and Letters 26.561-571.

_____.1942. Middle English WA in the speech of the Great Lakes region. American Speech XVII.226-234.

Phillips, Betty S. 1984. Word frequency and the actuation of sound change. Language 60.320-342.

Reed , Carroll E. 1977. Dialects of American English. University of Massachusetts Press.

Toon, Thomas E. 1978. Lexical diffusion in Old English. Papers from the Parasession on the Lexicon, 357-364.

_____.1983. The Politics of Early Old English Sound Change. New York: Academic Press.

Zipf, George. 1929. Relative frequency as a determinant of phonetic change. Harvard Studies in Classical Philology XL.1-95.

OSU WPL 34 (1986) 142-150

Introspection into a Stable Case of Variation in Finnish*

Riitta Välimaa-Blum

One of the more common signs of language change is vacillation between two competing forms. But once the change has run its course the older form typically disappears from the language. This, however, is not always the case: sometimes the archaic form remains as a stable variant of the newer, unmarked form due to, e.g., borrowing from spoken and/or written archaic dialects. In a study of language change in apparent time these coexisting alternants would have to be taken to signal ongoing change but this conclusion would be false since the vacillation is only seeming. The present study explores one instance of this kind of apparent vacillation not signalling change in Finnish.

When reading a chapter on the changes in Finnish phonology (Hakulinen 1979) I realized that many of the older forms are still (in some sense) in my vocabulary. According to Hakulinen some of these still exist in conservative dialects (presumably as the unmarked forms); some are marked as archaic but many have completely disappeared from modern Finnish. Below are listed some of the forms mentioned in Hakulinen which I do find in my stylistic repertoire, which naturally cannot be totally idiosyncratic but is necessarily shared:

	ARCHAIC		NEW	
(1)	kalatoin	~	kalaton	'without fish'
(2)	kankahat	~	kankaat	'moors'
(3)	saaliihin	~	saaliin	'of the prey'
(4)	tultihin	~	tultiin	'came' (impersonal)
(5)	urohon	~	uroon	'of the male'
(6)	harmaja	~	harmaa	'grey'
(7)	avajan	~	avaan	'I open'
(8)	tuleisi	~	tulisi	'would come'
(9)	meneisi	~	menisi	'would go'
(10)	menevi	~	menee	'he goes'
(11)	honkaan	~	honkain	'of the pine trees'
(12)	rukihin	~	rukiin	'of the rye'
(13)	tulkohon	~	tulkoon	'may it come'
(14)	hivus	~	hius	'hair'
(15)	munteilla	~	munilla	'on the eggs'
(16)	hevoinen	~	hevonen	'horse'
(17)	asja	~	asia	'errand; matter'
(18)	hipjä	~	hipiä	'skin'

The word pairs above were given in their historical order but, assuming that a naïve speaker at one particular point in time is not aware of the diachrony of his language, we could say that the second form in each pair is the basic synchronic form from which the first alternant is a deviation. For some of the 'deviations' we can write synchronic phonological rules by which these forms are derived from the new one. The form of the synchronic rule probably reflects the inverse of the

historical relation obtaining between the items in questions. For example, we could write a rule like

∅ ---> h/V__VC#

to give forms like (2 - 5). Or, a j-insertion rule for forms like (6) and (7). But for many of the words it is difficult to formulate rules since they appear to be isolated elements and no generalizations can be extracted from them. Thus the archaic part of my language contains archaic words per se, and generalizations by which archaic vocabulary items are created from the modern forms.

Not Vacillation but Borrowing

Do these archaic items exist in my speech because they are the older norm forms which are now disappearing? According to the source (Hakulinen), as was stated above, some of the forms are still used in conservative dialects; some are archaic, but some have completely disappeared. My dialect is not one of the conservative ones and none of the items above is part of this dialect in the usual neutral synchronic sense. They are never used in lieu of the unmarked forms in unmarked discourse, not by my generation, nor by any other generation in the same dialect group - there is no vacillation between two competing unmarked forms involving these items. Thus it seems that they do not exist in my speech because they are the disappearing older variants of the modern forms. What is their source then? I propose that the source is borrowing from conservative dialects (with which I am not intimately familiar) and from literary sources like the Kalevala, and other folk poetry, and folk songs.

The reason why the literary works mentioned above are seen as a possible source is that every Finn has to read at least parts of the Kalevala at school at some point or another; also other folk poetry, such as the Kanteletar, is studied, and folk songs are sung in music education classes. And, what is more important psychologically, these works are highly valued - they are some of the "national treasures" and as such they belong to the whole population. This fact makes them "common currency". Since everybody is exposed to samples of these works, they would be an obvious candidate for the origin of the archaisms under consideration.

Naturally, it is difficult to pin down exactly the source from which a lexical item has entered one's mental lexicon. I may have learned them, for example, from my schoolmates or relatives or mass media, etc., but the ultimate source of the old forms is proposed to be the archaic language of above mentioned and comparable works, and possibly conservative dialects.

Dowty has proposed that speakers are "potentially capable of remembering that they have heard a newly derived word for the first time, in a way that they very rarely recall hearing a sentence for the first time" and thus "speakers are able to distinguish between actual and merely possible (but well-formed) words in a way that they are not able to distinguish between actual and merely possible sentences" (1978: 120). These facts suggest to Dowty that lexically derived expressions

would be learned individually, unlike sentences. I do not agree with this for all lexical acquisition but it is certainly plausible in the present case. The deviant derivations are learned by persons who already are fluent speakers of Finnish. At this time they are, however, learning something that they know is deviant, and maybe for this reason some features of the source get attached to the set of connotations of the non-normal forms.

It is known that the writing system for a language can change language at the phonological and even the morphological level (Polomé 1981) but more relevant for the case on hand is that written material can function as a source of borrowing. Bloomfield (1933) cites examples of dialect borrowing of this kind which actually have led to change. Misunderstood archaic words may introduce new words or new meanings for words, e.g., derring do was 'daring to do' but became 'brave actions' (ibid., 487). But even some obsolete words may become reintroduced into the language from written sources: sooth and guise are examples of such words in English (ibid.).

Spelling pronunciation of English is one of the sources of variation and change comparable to literary borrowing: often is often pronounced with a [t]. Also completely new words come to language from writing: Comsomol, prof, and lab are such items. Bloomfield's examples make it indeed plausible that the ultimate source of these now archaic forms of Finnish could be the Kalevala and other similar written works.

The first page of the Kalevala gives us some of the forms mentioned above (the translation comes from an English version of the collection (1963)):

1.	Mieleni minun tekevi	It is my desire,
2.	Aivoni ajattelevi	it is my wish
3.	Lähteäni laulamahan,	to set out to sing,
4.	Saa'ani sanelemahan,	to begin to recite,
5.	Sukuvirttä suoltamahan	to let a song of our clan to glide on
6.	Lajivirttä laulamahan;	to sing a family lay.
7.	Sanat suussani sulavat,	The words are melting in my mouth,
8.	Puhe'et putoelevat,	utterances dropping out,
9.	Kielelleni kerkiävät,	coming to my tongue,
10.	Hampahilleni hajoovat.	being scattered about my teeth.

The first two lines give samples of type (10) given on the first page:

	OLD	NEW
(10)	menevi	~ menee

And lines (3) - (6) illustrate types (2) and (3):

(2) kankahat ~ kankaat
(3) saalihin ~ saaliin.

Line (8) has a form parallel to

(11) honkaan ~ honkain

Later in the Kalevala (1943, 44) we have

Sio silkillä hivusta! 'tie up your hair with silk ribbons'

where we have the archaic hivus as in (14).

Actually even whole phrases are used in the same stylistic contexts as the other forms:

Vaka vanha Väinämöinen 'Steadfast old Väinämöinen'

This phrase is repeated over and over again in the Kalevala. A thorough search would probably give citations of most of the older forms, if not in the Kalevala, then in some other archaic work which is still available and read today.

There are other, non-Kalevalaic words in the above mentioned chapter by Hakulinen (1978: ch. 2) which I use but for which the stereotype (to be discussed below) is somewhat fuzzy. It seems that while using these one is trying to evoke a foreigner image or an image of someone not quite mastering Finnish. These are, I believe, also borrowings from some unspecified written works. For form (20) below, for example, Hakulinen (p. 38) gives as one source a document from the 16th century (Agricola) samples of which are read during history classes. These forms are used stylistically like the other archaic forms and in that sense they have the same status as the Kalevalaic forms and therefore they are also problematic in the same way, as we will see below. Some of these other forms are the following:

	OLD	NEW	
(19)	vodottaa	~ odottaa	'to wait'
(20)	söö	~ syö	'he eats'
(21)	töö	~ työ	'work'
(22)	nyyt	~ nyt	'now'

Well-defined Stylistic Function and Value

Synchronically these older forms have an archaic and humorous value while the modern forms are unmarked. Thus they would seem to exemplify Kuryłowicz's fourth "law" of analogy: when two morphemes are undergoing differentiation the new form corresponds to the basic function while the older one has a derived, secondary function (1945). For example, the extension of the archaic [hevoinen] 'horse' is the same as that of the new [hevonen] but the intension is somewhat funny. The same holds for all of the pairs and thus the older elements are relegated to some special functions alone. The phenomenon is comparable to the one Searle refers to in his question "Knowest thou him who calleth himself Richard Nixon?" This question gets a different response than the following: "Do you know Richard Nixon?" (1975: 76). The hearer understands the intended meaning and responds accordingly, i.e., differently.

All of these archaic items belong to specific styles: mainly (in Joos's terms) to the intimate and casual styles. The use of these styles

involves intimate friends and/or members of a group equal in terms of age or social status, and more or less informal situations. Thus these archaisms occur only in limited, nonlinguistically definable contexts. But they are nevertheless widely shared since, for example, they can be heard on TV and radio shows. The function of their use is that of humor, e.g., to announce "let's keep the situation relaxed and informal", but their specific meaning or connotation has to do with something we might call Kalevalaic (excluding the "fuzzy" set (19) - (22) from Agricola's times); the stereotype being evoked has a Kalevalaic character. I do not wish to claim that each single item above has exactly the same stereotype for every user but that a significant subset of them is shared and this subset has a Kalevalaic flavor.

Psychologically there is a difference in the stereotypes between the archaic forms and those of some synchronic dialect forms. For example, words like (23) -(28) are at present unmarked forms in dialects other than mine.

(23)	mie, mää	= minä	'I'
(23)	kolome	= kolme	'three'
(24)	palakka	= palkka	'salary'
(25)	syärä	= syödä	'to eat'
(26)	tehrä	= tehdä	'to do'
(27)	hyö	= he	'they'
(28)	hää	= hän	'she/he'

The use of these evokes different stereotypes than the archaic ones. In this case the character is clearly synchronic and actually even geographically circumscribed. Thus in both cases the forms are used for stereotyping but the images being evoked are different. And this would fit neatly with the ideas of Dowty's mentioned above: the initial context of learning is remembered and consequently it has become part of the derivation.

Scott (1982) has discussed what she calls 'vivid language', and its role in language change. Vivid language, according to her, is characterized by intentional but interpretable deviance, and novelty. By this definition the archaic forms clearly count as vivid language: they are intentionally deviant and also interpretable, and they have novelty value, at least in particular contexts. Scott gives examples of how vivid language may become conventionalized and result in changes in a language. Thus vivid language is relevant to historical studies and therefore vivid language, like different styles, must be part of the corpus in diachronic explorations. I will return to this point below.

Unlikely Future Kyriolexis

If the archaic forms under consideration indeed have become conventionalized, and since they exist alongside the new forms, is it to be expected that they become one day the kyriolexis (Householder 1983) or the modern unmarked forms? Householder dicusses how a speaker has a subconscious norm (or as I would call it, a pragmatic template) against which lexical items are matched when they are heard or used. One item typically is the norm or kyriolexis (16. hevonen) while the other (16. hevoinen) is the tolerated deviation (cf. Kuryłowicz and Scott above),

and when the tolerated deviation is "promoted" it "overthrows" the previous norm and becomes the new kyriolexia.

Since these archaic forms are stylistic variants, i.e., vivid language, and their use is nonlinguistically governed, and they are not the kyriolexia, is it to be expected that they will become the new norms? It would not be impossible, but in the present case this would seem very unlikely because of their well-defined stylistic function, strong archaic flavor and humorous value. They are not in competition with the new forms but have a clearly assigned place in the language use. But on the other hand, they are not likely to disappear either, partly due to their above-mentioned function, but also because their proposed sources are continuously available and they enjoy some kind of prestige.

In-group Prestige

It is commonly assumed that the source of borrowing possesses prestige, but as Labov has made it clear, prestige is not a unidimensional concept. He stresses that language change in general and thus also borrowing cannot be understood apart from the social setting where it takes place (1978:23). Social meanings thus get attached to certain alternative forms, and, to quote Sturtevant, these "rivals shall acquire some sort of prestige" (ibid.: 3). But the crucial point is that the prestige lies in the eye of the beholder - if someone is perceived as possessing some desirable feature(s) he may well be imitated regardless of his actual social standing.

How would prestige explain the present case of borrowing? As stated above the Kalevalaic works are highly valued by the Finnish people in general; moreover, there is some degree of identification with the Kalevala in the historical sense, for the Finns sometimes address themselves as the Kalevalaic people:

"Mehän ollaan tämmöstä Kaleva/kalevalaista kansaa!"
we are this Kaleva's/Kalevalaic people
"We are, after all, this sort of Kalevalaic people"

The identification is in the sense of continuity, in the sense of sharing something that forms one of the roots of "Finnishness". Even if I speak here of introspection, the forms are shared, as was pointed out above, so that they can even be heard on TV and radio shows. Maybe there is something similar going on here as was on Martha's Vineyard (Labov: 1978): these archaic forms with their Kalevalaic stereotypes are shared linguistic conventions, and they are used to convey some kind of humorous in-group sentiment. Their use is stylistically governed but they are part of the stable linguistic repertoire; the forms are used only with "insiders". But Finland is small and culturally relatively homogeneous, and all the Finns are apparently seen to be insiders.

Relevant Data from Styles

Hammarström suggests that "in synchrony, neither "older" nor "younger" forms...nor change can be considered, as they are not part of that knowledge which is used in communication" (1982: 51). For him the

use of older forms of a language, or of foreign words in one's speech, is more like a case of mention rather than of use, and thus must be eliminated from the synchronic description of a language. As has been shown above the archaic forms are used in "true" communication, in true social discourse; they are not instances of mention at all but carry their own intended, situationally relevant meanings.

Buccellati notes that too often anything "formally bizarre" or unexplainable is assigned to stylistics and thus discarded from the data (1981). For Buccellati, style is not a synchronic phenomenon either, but exclusively a diachronic "distributional category": "a recurrent selection of relatively idiosyncratic features" (1981: 808). In linguistics these features can be morphological, syntactic or lexical items "sufficiently distinctive to acquire stylistic value" (ibid.: 809). Idiosyncracy in this definition refers to the distinctiveness of the items within a given "assemblage" of language as opposed to another "assemblage". Here styles can be seen as particular assemblages and the archaic items have distinctive value in some styles since they are not found in all styles.

I would agree with Buccellati, and Scott above, and consider stylistic choices as 'vivid language' relevant for diachronic analysis, but contrary to both Buccellati and Hammarström, also for synchronic analysis. After all, many parts of language in general could be allocated to certain styles alone, but we would not discard those parts. The use of slang and jargon, for example, is also stylistically governed. If archaic items are to be eliminated from linguistic analysis, so too, then, are jargon and slang.

Problems for Diachronic Studies in Apparent Time

Now that the archaic forms are seen as part of the relevant corpus, their retention in the data may pose problems for a study of language change in apparent time. In this kind of study we need real time evidence to show that the part of the data being studied has been the unmarked data at some earlier time (Labov 1978: 275). We do have this evidence for all of the forms: they have been the neutral forms at some earlier point in time. But if a linguist now in his apparent-time study of Finnish finds these archaic forms alongside the unmarked forms, he would not have any motivation for saying that these forms are not in competition with the unmarked forms. Consequently he would be forced to conclude that he is facing the typical vacillation situation of language change and that Finnish is undergoing change. Additionally, he might not find these elements in the speech of the very young population which has not yet been exposed to these writings and this fact would only support his conclusion. But this conclusion conflicts with the facts: most of these archaic items are synchronically living, stylistic variants of the unmarked synchronic forms. They exemplify the fact that a native speaker of a language knows a considerable range of facts about this language. The sources may be old texts, ordinary school books, mass media, home and friends, etc. We cannot forget in historical or synchronic studies that our language is a multidimensional entity in terms of styles and registers and speakers have mastery of this entity.

Summary

We have examined the use of some archaic lexical items of Finnish everyday discourse. These items occur only in specific, nonlinguistically describable contexts. They involve a Kalevalaic, archaic stereotype and they probably are instances of dialect borrowing, partly from literary sources, possibly also from conservative dialects. Since they are part of the stable linguistic system, they are valid data for both synchronic and diachronic analyses. Their coexistence, however, gives the wrong impression that the speakers are alternating between two competing forms. But this is only apparent vacillation; actually the phenomenon is something resembling a "Kalevalaic diglossia". Thus not all variation between archaic and unmarked forms is vacillation signalling change in progress.

* This paper was read at the annual meeting of the Finno-Ugric Studies Association of Canada in Montréal in May 1985; the theme of the meeting was the 150th anniversary of the first publication of the Kalevala.

References

Bloomfield, Leonard. (1933). Language. New York: Holt, Rinehart, and Winston.

Buccellati, Giorgio. (1981). Principles of Stylistic Analysis, in Yoel L. Arbeitman & Allan R. Bomhard (eds.). Bono Homini Donum: Essays in Historical Linguistics in Memory of J. Alexander Kerns, II. Amsterdam: John Benjamins B.V., 807-836.

Dowty, David. (1978). Applying Montague's Views on Linguistic Metatheory to the Structure of the Lexicon, CLS Parasession on Lexicon, 97-138.

Hakulinen, Lauri. (1979). Suomen kielen rakenne ja kehitys (The Structure and Development of Finnish). Helsinki: Otava.

Hammarström, Göran. (1982). Diachrony in Synchrony, in Peter J. Maher, Allan R. Bomhard and E. F. Koerner (eds.). Papers from the 3rd International Conference on Historical Linguistics. Amsterdam: John Benjamins B.V., 51-64.

Householder, F. W. (1983). Kyriolexia and Language Change, Language, 59.1-17.

Kuryłowicz, J. (1945). La nature des proces dits 'analogiques' , Acta Linguistica 5, 15-37.

Kalevala, (1943), Porvoo: WSOY.

The Kalevala or Poems of the Kalevala District, (1963). transl. Francis Peabody Magoun, Jr., Cambridge, Mass.: Harvard University Press.

Labov, William. (1978). Sociolinguistic Patterns. Philadelphia: University of Pennsylvania Press.

Polomé, Edgar. (1981). Can Graphemic Change Cause Phonemic Change, in Yoel L. Arbeitman and Allan R. Bomhard (eds.), Bono Homini Donum: Essays in Historical Linguistics in Memory of J. Alexander Kerns, II. Amsterdam: John Benjamins B. V., 881-888.

Scott, Janet. (1982). Vivid Language and Language Change, in Anders Ahlqvist (ed.). Papers from the 5th International Conference on Historical Linguistics. Amsterdam: John Benjamins Publishing Company, 304-315.

Searle, John. (1975). Indirect Speech Acts, in Peter Cole and Jerry L. Morgan (eds.). Syntax and Semantics. Speech Acts. New York: Academic Press.

OSU WPL 34 (1986) 151-164

Diglossia in Ancient India

Gina M. Lee

1. *Introduction*

The rich variety of languages spoken in Modern India, with representatives of several language families (Indo-Aryan, Dravidian, and Munda, as well as English) has sparked much interest in the ramifications of language contact in India, and South Asia in general. In particular, the relationship of some Indian languages spoken within the same speech communities has been said to be diglossic: Gair (1968) and De Silva (1974) have proposed that the relationship between the literary and colloquial varieties of modern Sinhalese (spoken in Ceylon) is diglossic.

Like its present day counterpart, ancient India was a multilingual area. Not only were the ancestors of modern Indo-Aryan languages (namely Sanskrit and the Prākrits) spoken in the same region, but also the forerunners of modern Tamil and Munda. Diachronically speaking, Sanskrit (both Vedic and Classical) is considered Old Indo-Aryan, and the Prākrits are traditionally considered Middle Indo-Aryan. But many (e.g. Emeneau 1966) have noted that Sanskrit and Prākrit were also spoken during the same time period.

Although Indo-Aryan scholars have continually referred to the Prākrits as the popular dialects and to Sanskrit as the language of the learned, the possibility of diglossia existing in ancient India was not discussed in depth until Hock and Pandharipande (1976).[1] Even so, later scholars have not expanded on the hypothesis of diglossia during ancient times; Deshpande (1979) discusses instances of conflicting sociolinguistic attitudes in ancient India, but does not provide direct evidence for or against diglossia.

The purpose of this paper is to evaluate the evidence (presented primarily by Hock and Pandharipande) in favor of a diglossic relationship involving Sanskrit and the Prākrits in terms of Ferguson's original 1959 definition.[2] In making their claim that Sanskrit and Prākrit were used in diglossic situations as early as the time of the Rig Veda, Hock and Pandharipande give three types of evidence. They present as the best-known evidence the language differentiation in the Sanskrit drama, in which Sanskrit was used by characters representing the higher castes and various Prākrits were used by characters representing the lower social castes.

Also cited as evidence are various passages from the primary Sanskrit literature, most notably from the writings of the grammarian Patañjali (c. 150 B.C.). Patañjali notes in referring to Pāṇini (1.1.1., 259:13) that there are differences between the śiṣṭabhāṣā, the language of the learned, and the lokabhāṣā, the language of the common people. The Nāṭyaśāstra, the oldest treatise on Sanskrit drama (attributed to Bharata, c. third century A.D.), gives factors which determine whether a character may or may not use

Sanskrit. The important factors were social status, caste, occupation, and social context. The level of education was an important factor, for well-educated people were to use Sanskrit.

Hock and Pandharipande mention briefly, as a third type of evidence, the occurrence of so-called *hyper-Sanskritisms*, hypercorrections of Sanskrit (or, in most cases, Prākrit) forms which are intended to avoid patterns found in Prākrit.

The evidence supplied by the ancient Indian commentators and grammarians provides strong evidence for, if not diglossia, at least some level of conflict between the two language varieties. Such evidence will not be disputed here. What is open to question, though, is the evidence from the Sanskrit drama of the period 100-1000 A.D. It appears that the drama may not be a reflection of the actual structure of ancient Indian society. On the other hand, though, the evidence provided by hyper-Sanskritisms can be shown to be more important to the argument for diglossia than what Hock and Pandharipande claim. This paper contains the results of a systematic investigation of hyper-Sanskritisms.

2. *The Evidence from the Drama*

The Sanskrit drama provides evidence for, at the very least, the literary coexistence of Sanskrit and Prākrit. In general, Sanskrit was used by characters of the higher social castes; within the same play, various types of Prākrits were used by characters of lower social groups, which included comic characters and women. According to Rājaśekhara (c. 900 A.D.), a dramatist who had a special interest in language, Prākrit is "smooth" (hence, its general use by women) while Sanskrit is "harsh" (hence, its general use by men.) Although the Nāṭyaśāstra gave elaborate rules for the use of language in the drama, such rules were by no means rigid. A considerable amount of variability existed, particularly in the use of Prākrit.

Authority figures such as kings and generals were to use Sanskrit; and as might be expected, Brahmins were also to use Sanskrit. Some female characters used Sanskrit: the chief queen, the ministers' daughters, and occasionally Buddhist nuns, female entertainers, women artists, and allegorical female characters. Without fail the descriptions of battles, peace negotiations, and omens required the use of Sanskrit.

On the other hand, the Prākrits were used by women other than those mentioned above, as well as by men of lower rank. Particular dialects were ascribed to particular types of people, although the use of a particular dialect differed from author to author. Śaurasenī was generally used by women of "good family", their servants, and middle class males. Māgadhī, another well-known Prākrit, was used by men living within the women's apartments, diggers of underground passages, bartenders, and, interestingly,

by the hero in times of danger (possibly expressing his "feminine", emotional side). Gamblers used Avantī and Dākṣinātyā. Sometimes, two varieties of Prākrit were used within the same play: Kālidāsa (c. 400 A. D.) used Śaurasenī in prose, Māhārāṣtrī in verses.

However, according to one of the earliest scholars on Sanskrit drama, Sylvain Lévi (Le théâtre indien, 1890), the drama could not have reflected a diglossic situation. The plays, in his view, were originally composed in Prākrit. As a result of the rise of Sanskrit as the language of literature as well as religion, the drama developed a mixture of the two varieties. Moreover, Lévi argued that "India . . . was never anxious for contact with reality, and it is absurd to suppose that the mixture of languages was adopted as a representation of the actual speech usage of the time . . . " (quoted in Keith 1924: 46).

But the evidence so far is that the drama was not secular in origin, but religious, arising from epic recitations. Moreover, in the work of the earliest known Sanskrit dramatist, Aśvaghoṣa (c. first century A.D.), Prākrit appeared mainly in the dialogue, while Sanskrit appeared mainly in the verses. Thus, it appears that in the early dramas, Prākrit was introduced into what was essentially a Sanskrit drama, in order to reflect the status of the inferior characters.

Other arguments can be made that the language usage in the drama cannot be due simply to an imitation of the real life situation. The Prākrits of the later dramas were in some respect different from the Prākrits spoken in everyday situations. As early as 400 A.D., the Prākrits used in the drama began to take on artificial, literary forms. Reference is made to vibhāṣās, stereotyped variants of the "more normal" Prākrits, which refer to some literary Prākrits. For example, people of menial occupations used certain Prākrits: herdsmen used Śābarī or Ābhīrī; charcoal burners, hunters, and carpenters also used Śābarī. (But the existence of literary forms does not necessarily mean that the Prākrits used in the drama are completely unreliable as evidence; in a study of Irish literary dialects Sullivan (1980) argues that literary dialects can reflect characteristics of the actual speech.)

Moreover, there is evidence that the drama appealed to only a limited Indian audience and was intended to be viewed only by members of the higher social classes. As early as 900 A.D., chāyās, translations of the Prākrit portions into Sanskrit, were common. No evidence exists for translations of the Sanskrit portions into Prākrit, which suggests that the dramas were written mainly to be viewed by those who knew Sanskrit, i.e. the learned. Keith (1924: 242, 369-371) argued that the Sanskrit playwright's works were aimed mainly at the learned. Using (in part) information from unpublished texts, Balbir stated that ". . . the Sanskrit drama perhaps was never a light amusement of everyday life . . . it is obvious that the Sanskrit drama was intended to be a drama of the elite, enjoyed by qualified persons . . . a refined product religiously presented as an offering before a discriminating audience . . ." (1962: 44)

The drama could only be appreciated by a special group of people who were not only trained to appreciate the aesthetic qualities of the drama, but who could also be empathetic with the characters on stage. Appreciation for the drama could only be cultivated by a certain amount of study. The ideal spectator had to be knowledgable about many things, among them the "rules of dialects . . . (and also) grammar" (Balbir, quoting from the Nāṭyaśāstra). Citing Lévi, Balbir states that "all the spectators are not apt to relish the rasa ['taste, feeling']; it is a sort of prize one has to deserve after an assiduous study of poems and healthy and delicate impression accumulated from the previous births." These people are referred to by various terms in the primary literature: as prekṣaka, sāmājika, sabhya, and sabhāsada.

Sabhāsada refers to 'an assistant at a meeting or assessor in a court of justice.' Prekṣaka means 'looking at, viewing or intending to view', as well as 'spectator, member of an audience'; but it could also have the meaning of 'considering' or 'judging'. Sāmājika is a term that was neutral in meaning, meaning 'spectator, member of or assistant at an assembly'. Sabhya could be neutral in meaning as well, meaning 'being in an assembly hall or meeting room, belonging to or fit for an assembly or court'; it could also, however, mean 'suitable to good society, courteous, polite, refined, civilized, not vulgar, decorous' (as speech); or 'a person of honorable parentage'. Such spectators were, for the most part, members of the higher social classes.

It was essential that audience members be well-qualified to view the Sanskrit drama, for the audience members decided whether the play was a hit or not. Every ancient Indian audience had a sabhāpati (literally 'audience-ruler'), the guest of honor, who made the final decision as to the success of the play. The sabhāpati had advisors to guide him in his decision; each advisor was a specialist on a particular aspect of drama. Also present at the Sanskrit drama were "assessors", people of various occupations whose job was to evaluate the acting of individual performers. What is of interest here is that grammarians were also present as assessors.

The common folk also attended dramas; their opinions on the success of the play were acknowledged, but were not respected. According to the Nāṭyaśāstra, the audience was divided into two types: divine and human. The divine refers to the "cultured audience who generally take interest in deeper and more subtle aspects of a dramatic and as such are above ordinary human beings" (Balbir quoting translation from Ghosh, p. 513, fn. 17 & 15). The human element refers to the common people who were appreciative only of superficial aspects of the drama, and not of the deeper aspects.

Certainly the Sanskrit drama was something that was staged only on special occasions, such as military victories, festivals honoring the gods, or weddings. The playhouses (the nāṭyaveśma, nāṭyagṛha, and prekṣāgṛha) are described in the literature as having elaborate seating arrangements, with the best seat in the house given to the sabhāpati. In some instances, they are referred to as "palace-theatres", which may indicate that some plays were staged within makeshift theatres within the royal palaces.

Given that the Sanskrit drama was viewed by a limited audience, the bilingual nature of the Sanskrit drama does not provide conclusive evidence for diglossia. Stronger evidence for the high social status assigned to Sanskrit comes from hyper-Sanskritisms.

3. *The Evidence from Hyper-Sanskritisms*

Linguists have devoted a fair amount of attention to hypercorrections, the use of a form based on attempts to avoid forms found in low prestige dialects. DeCamp (1972) mentions various examples of phonological hypercorrection in American English, such as /r/ insertion in some San Francisco dialects, or Jamaican Creole substitution of /ɸ/ for /t/ in words such as /fIlɸr/. In such forms, there is an effort, conscious or not, to avoid using forms which are phonologically similar to low prestige forms, even if they are not low prestige pronunciations. /r/ insertion appears to have originated from an attempt to avoid using what could appear as /r/ deletion; the /ɸ/ for /t/ substitution resulted from an awareness of the converse substitution in low prestige dialects. In addition, Labov (1972) describes hypercorrections in terms of the frequency of usage of correct forms; the middle class is likely to use prescriptively correct forms more often than higher social classes.

In this discussion, I am using the term *hypercorrection* in a more general sense than what has been traditionally used: to refer to any morphological change which originates as an attempt to avoid using forms which contain phonological patterns found in a low prestige dialect. Since the original forms do not violate phonotactic (or syntactic) rules, such "corrections" are unnecessary from a structural viewpoint; hence, they are *hyper*-corrections. Traditionally, hypercorrections have been used to refer to prescriptively/etymologically incorrect forms which originate in such manner, but prescriptive or etymological correctness/incorrectness is unimportant. What is important is the social forces behind such modifications.

Perhaps the best examples of such forms found in a language not usually considered a living language are found in the hyper-Sanskritisms, phonological hypercorrections (limited to certain lexical items) which originated as modifications of Prākrit forms, or of Sanskrit forms which contain patterns found in Prākrit. Some, if not all, Sanskrit speakers must have been aware of the phonological differences between Sanskrit and the Prākrits. In a few instances, Prākrit words which are borrowed into Sanskrit are modified to sound more Sanskritic. For example, Sanskrit has a noun utkuruṭa- 'dumthemp', which originates as a hypercorrection from the Prākrit form having the same meaning, ukkurudi-. The Prākrit reflex of Sanskrit tk is kk. From a phonological standpoint, there is no motivation to change the kk sequence to tk because kk can occur in Sanskrit, as in Skt. kakkola- 'a species of plant'. The only motivation for such a change, if not due to loan phonology, is a social one: Sanskrit speakers wanted to avoid using the kk sequence which, in principle, could be perceived as a Prākrit sequence.[3]

In some cases, words which were Sanskrit in origin were erroneously perceived as Prākrit and modified so as to be "more Sanskrit". Utsuka- is a modification of Sanskrit *ucchuka-, which comes from Old Indic *icchuka-. Since the cch sequence in *ucchuka- is identical to the cch sequence which is the Prākrit reflex of Sanskrit ts, the Sanskrit form changed in a direction away from (what was perceived as) Prākrit.

I examined all cases of hyper-Sanskritisms (primarily) from two sources. One of the earliest works which refers to hyper-Sanskritisms (and uses the term *hyper-Sanskritism*) is Bloomfield and Edgerton's work on Vedic phonetics (1932: 20). The influence of Prākrit on Sanskrit is manifested in two ways: first, by Prākritisms, changes in Sanskrit forms in the direction of Prākrit. Many writings in Vedic Sanskrit (including the Rig Veda) contained unusual Sanskrit forms which are phonetic variants that follow sound patterns in Prākrit. For example, the form tvaṣṭr- 'creator' has a variant form tvaṣṭri-, which appears to be influenced by the occurrence in some Prākrits of ri (or ru) for Sanskrit ṛ. Secondly, the opposite may happen: the Sanskrit form may have a variant form which is modified in a direction away from Prākrit-like forms, or toward a variety of Sanskrit which cannot be perceived as having any Prākrit influences, as in the hyper-Sanskritisms. The hyper-Sanskritisms cited in Bloomfield and Edgerton appear to be hypercorrected forms of Sanskrit forms erroneously perceived as Prākrit. It is these types of hyper-Sanskritisms which Hock and Pandharipande cite as evidence for diglossia.

Mayrhofer (1966) takes a different approach to hyper-Sanskritisms. He defines a *Hypersanskritismus* in the following way:

> Perhaps still more frequently than the undertaking of the pure or almost unchanged dialectal forms was also the case that these have been again adapted falsely to the high dialect. . . . In several cases . . . we encounter strange *Rück-Sanskritisierungen* of such Middle Indic (or, even only to be regarded as M[iddle] I[ndic], in truth correct Old Indic) words and these *Rückbildungen* are again a fact, which the Old Indic etymology by all means has included. (my translation of Mayrhofer 1966: 9)

In volume I of Mayrhofer's work, I examined each entry to see whether it could be attributable to a hyper-Sanskritization.[4] (Unfortunately) Mayrhofer uses five terms to refer to such hypercorrections: *Hypersanskritismus*, *Rück-Sanskritisierung*, *Rückbildung*, *falsche Sanskritisierung*, and (occasionally) *Sanskritisierung*. These are distinguished from Prākritisms (*[ein] Prākritismus* or *dialektische Formen*).
Mayrhofer is mainly concerned with modifications in Prākrit forms which eliminate certain patterns found in Prākrit. However, such modifications are, from a social standpoint, the same type of modifications that occur in Bloomfield and Edgerton's hyper-Sanskritisms.

In examining hyper-Sanskritisms, I found that they are not limited to only one or two categories, but that there were apparently many *types* of hyper-Sanskritisms that took place. This has two important implications. First, the occurrence of such types suggests that hypercorrection may play a greater role in morphological change than previously thought. Many have acknowledged that language change can arise as a result of speakers' tendency to regularize, as in analogical change. Occasionally morphological changes occur which involve apparent reversals of established sound correspondences. The best explanation for such reversals, especially in situations involving literary and colloquial variants, is hypercorrection. Thus, social factors can play an important role in accounting for changes in the phonological shape of words.[5]

On the basis of the similarity in the types of forms found, as well as the variety of types, it appears that hyper-Sanskritisms are not a "grab-bag" group of words whose phonetic shape cannot be explained, but rather are words which reflect an actual sociolinguistic phenomenon in ancient India. It could not simply be a coincidence that all of the patterns found involved a change from (apparent) "Prākrit" to "Sanskrit"; the only possible motivation for such changes is hypercorrection. Though there is no semantic pattern in these forms, the hyper-Sanskritisms fall into a set of distinct groups, as follows:[6] [note: unless otherwise indicated, the original forms are Middle Indic; forms which are indicated as variants come from original Vedic forms; MI = Middle Indic, OI = Old Indic.]

Modifications of Consonant Sequences:

a. One of the Prākrit reflexes of Sanskrit ts is c(h) (frequently doubled to cch), as in Skt. matsara-, Pkt. macchara- 'cheerful; intoxicating'. cch is a possible (and common) word-internal sequence in Sanskrit, as in gaccha- 'tree'. A number of hyper-Sanskritisms were found involving ts for c(h)/cch:

gutsa- from gaccha- 'bundle'
utsuka- from *ucchuka-, OI icchu-, 'restless, anxious, longing for'
utsādana- from ucchādana- 'rubbing'
kudyamtaī-/kudyamtaya- from *kudamac(h)- 'house lizard'
jugupsa- 'avoids, detests' from MI *jugucchu- (Pāli jigucchā-), 'abhorrence'; desiderative of gup- 'protect'.

b. Prākrit occasionally has (k)kh for Sanskrit ks, as in Skt. bhikṣu-, Pkt. bhikkhu- 'monk'. In Apabhraṃsa, such a change occurs regularly, as in Skt. kṣatriya-, Apaṃ. khattiu- 'warrior'. (k)kh was possible in Sanskrit, as in khakkhati '(a)he laughs'. Nonetheless, Sanskrit speakers substituted kṣ for (k)kh in some words:

akṣauhiṇī- from MI *akkhohiṇī-, Pāli akkhobhaṇī- 'complete army'
kṣētra- from khētra- 'breach, tunnel'
kṣīv- from khiv- 'spits'
rukṣa- from MI rukkha-, OI vṛkṣa- 'tree'
kṣvel- from khel- 'leap, jump, play'

c. The Prākrit reflex of Sanskrit tk is kk, as in Skt. utkara-, Pkt. ukkero, 'heap.' kk is a possible Sanskrit sequence, as in kakkola- 'a species of plant'. I found one example of a semantically related hypercorrected form: utkuruṭa- for ukkurudī- 'dustheap'; also, muktā- from *muttā-, Pāli, Pkt. muttā-, OI mūrtā- 'pearl'.

d. Prākrit kk can also arise from Sanskrit rk, as in Pkt. akka-, Sanskrit arka- 'ray, flash of lightning; sun.' In one hyper-Sanskritism, rk is substituted for kk: kurkuṭa- from older, literary kukkuṭa- 'cock.'

e. In some hyper-Sanskritisms, tt became st(h): kandoṭṭha- from Pkt. kaṇḍotta-, kaṇḍutta-, OI kaṇḍata- 'blue lotus'; addhyuṣṭa- from MI addhutta-, OI ardhacaturtha- 'three and one-half'. Numerous examples of tt occur in Sanskrit: aṭṭa- 'watch-tower; market.' ṣṭh did not regularly become tt in Prākrit, but compare ṣṭ > tth in forms such as Skt. dṛṣṭi-, Pkt. diṭṭhi- 'sight'.

f. Prākrit shows bbh for Sanskrit dbh, as in Skt. sadbhāva-, Pkt. sabbhāva- 'good nature.' One type of hyper-Sanskritism involved dbh for bbh: adbhis/adbhyas (instr/dat, abl pl. of ap-) from *abbhis, abbhyas 'water'. A compound form abbhakkha- 'living upon water' shows that bbh can occur in Sanskrit.

g. Sanskrit rv became vv in Prākrit, as in Skt. sarva-, Pkt. savva-, 'all.' vv became rv in hyper-Sanskritisms:?

urvarita- from uvvariā-, 'left, left over'.
carv- for OI *cavv-, 'grinds with the teeth, chews'

h. In one hyper-Sanskritism, rg comes from gg, as in argala- from MI aggala-, OI *agra-la- 'going beyond'. gg is a possible sequence in Sanskrit, as in the compound diggaja- 'one of the elephants in the four quarters (who support the earth)'.

Modifications of Individual Consonants:

i. Dialectally in Prākrit, ḍ was substituted for ṭ in a limited set of words, all of which are forms of the second person singular pronoun, e.g. ḍāva for tāvat 'your.' In some words, Prākrit shows ḍ for Sanskrit ṭ, as in Skt. parita-, Pkt. parido- 'around.' A few Prākritisms involved the interchanging of voiced stops for voiceless stops, as in the case of edagvē-

for etagvā- 'of variegated color, shining (horses)', and piga- for pika- 'cuckoo'. Hypercorrections in the other direction occur as well: devi- '(perhaps) nursery term for mother' has the variant form tevi-.

j. Some Prākritisms show y for y, as in jaṇovāda- from OI janāpavāda- 'gossip, ill report'. The converse hyper-Sanskritism occurs: kapāṭa- from kavāṭa- 'leaf of a door.'

k. Prākrit regularly shows h where Sanskrit has aspirated stops, as in the following: Skt. sukha-, Pkt. suha- 'pleasure'; Skt. laghuka-, Pkt. lahua- 'small one'; Skt. pathika-, Pkt. pahia- 'traveller'; Skt. nidhi-, Pkt. nihi- 'treasure'; and Skt. abhinava-, Pkt. ahiṇava- 'fresh'. Some Prākritic forms showed a substitution of h for aspirated stops, as in kakuha- from kakubha- 'high, eminent, great'; gahaṇa- from gambha- 'deep'. The corresponding hyper-Sanskritism of dh from h occurs:

gudhera- from guhera- 'protecting'
avadhamsa- from Pkt. ohaṃso-, OI *avagharṣa- 'red sandal'

l. In some Vedic forms, j was substituted for original d (especially before y): dyut- has the variant jyut- 'shine'; original daha, imperative of han- 'strike, kill', became jahi. The opposite hypercorrection occurs as well: jyā- 'bow string', has the variant dyā-.

Vocalic Hyper-Sanskritisms:

m. Prākrit frequently reduced word final -as (-aḥ) to -o, as in Skt. drumas, Pkt. dumo, 'tree.' Final -o occurred in Sanskrit as a result of a sandhi rule involving the change of final -as to -o before voiced consonants, as in devo gacchati (from underlying devas gacchati). As might be expected, as is substituted for o in hypercorrections:

amas for OI *amo 'this', nom. sg.
adas for OI *ado 'that', nom. sg.

n. Prākrit occasionally shows i (and sometimes u) in place of vocalic r, as in Skt. dṛdha-, Pkt. dadha- 'firm'; Skt. amṛta-, Pkt. amia- 'nectar' and Skt. prākṛta-, Pkt. pāua- 'Prakrit'. Numerous Prākritisms show a substitution of i for vocalic r as in ghinnate from OI *gṛbhnāti, third singular present of grabh- (grāh-), 'takes'. Some hyper-Sanskritisms have r interchanged with i or u:

krcchra- from *kicchra- 'evil, bad'
masiṇa- from MI masiṇa-, OI mṛtsna- 'soft, mild'
(gotra)bhrd, a variant of gotrabhid 'opening the cow-pens of the sky' (of Indra and Brhaspati's vehicle)
jaivātrka- from OI *jaivatu-ka- (vrddhi of jīvātu-), 'long-lived'

rccharē- 'courtesan' from accharē- 'Apsaras' (name of female divinity)
ghusṛṇa- from Pkt. ghusiṇa- 'saffron'

o. The Prākrit dialects occasionally had forms with ru or ri for Vedic r (paralleling the modern pronunciation), as in bhrumi- for bhṛmi- 'whirlwind'. There is at least one hyper-Sanskritism corresponding to this: prsvē- occurs as a variant of prusvē- 'drop of water, rime, ice'.

4. *Conclusion*

The issue is whether Sanskrit and the Prākrits, which were used by speakers within the same speech communities, could be accorded diglossic status. Certainly there is evidence for the existence of a high and low variety, with Sanskrit holding the position of high prestige and Prākrit, low prestige, as Hock and Pandharipande argue. But the evidence from the Sanskrit drama does not conclusively prove the existence of diglossia, since the drama was written mainly for audience members who were essentially the upper crust of society and, as a possible consequence, did not accurately portray actual language usage. It is also likely that the use of Sanskrit and Prākrit in the drama, especially in the later works, was merely a matter of literary tradition, rather than a depiction of the real-life situation (Burrow 1973: 60; cf. also the occasional stereotyped use of Southern accents for inferior characters in American English).

The hyper-Sanskritisms, however, are stronger evidence for diglossia. Since neither the absolute number of hyper-Sanskritisms found nor the absolute number of hypercorrection patterns provide conclusive evidence for diglossia, my intention is not to provide a statistical argument for diglossia. There is no "magic number" of hypercorrected forms or patterns which conclusively indicates that speakers viewed each variety as having different social status. Moreover, the number of hyper-Sanskritisms found in Mayrhofer's dictionary does not provide a figure for the token frequency of *words* which underwent such hypercorrection. Some forms occur more frequently than others. In addition, the existence of hypercorrections in itself does not signal diglossia, since hypercorrections (of both phonological and morpho-syntactic nature) occur in non-diglossic situations, such as American English.

But if hypercorrection played only a minor role in accounting for morphological change within a language, then one would not expect to find many different types of hypercorrection. Certainly the occurrence of only one or two patterns could not be used as evidence for differing social attitudes toward the dialects. The large variety of hyper-Sanskritisms, with numerous different patterns, strongly suggests that there were conscious efforts on the part of Sanskrit speakers to avoid using forms which sounded Prākritic.

It appears that the Prākrits were not simply the dialects used by the *populi*, but were varieties that had low social standing. Sanskrit was, in

addition to being the language used by the learned, a variety that held much greater prestige than the Prākrits. Thus, in much the same way in which ancient Indian society was stratified, Sanskrit and the Prākrits were also socially differentiated.

Notes

My thanks to Brian Joseph for his comments on earlier versions of this paper.

1. De Silva (61-62) argues that, as early as 800 B.C., Vedic and Classical Sanskrit were used diglossically, with the Classical language as the high variety and Vedic as the low variety.

2. Ferguson's definition of diglossia, in its entirety, is as follows:

> . . . a relatively stable language situation in which, in addition to the primary dialects of the language (which may include a standard or regional standards), there is a very divergent, highly codified (often grammatically more complex) superposed variety, the vehicle of a large and respected body of literature, either of an earlier period or in another speech community, which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any sector of the community for ordinary conversation. (1959: 336)

In this discussion, I am using a simplified version of Ferguson's "classic" definition: namely, situations involving a variety which is assigned high social status, while the other variety is regarded as having low status by speakers. This simplified version appears to be the sole criterion used by Hock and Pandharipande in their analysis (113); they do not discuss criteria other than prestige. The criterion of function, with mutually exclusive tasks assigned to each variety, is a natural consequence of the occurrence of high and low varieties.

Also, this simplified version represents the essence of Ferguson's definition, which distinguishes diglossic situations from cases involving regional and stylistic variation. The two varieties must have a moderate amount of divergence, in the sense that they must be different enough so as not to be styles, but they must be similar enough so as not to be unrelated languages. Ferguson's definition differs significantly from Fishman's (1972) and Fasold's (1984) later modifications. Fishman agrees with Gumperz's argument (1961, 1962, 1964a, 1964b, 1966) that diglossia involves two functionally differentiated language varieties of *any type*, regardless of their degree of divergency. According to Fishman, "diglossia is a characterization of the social allocation of functions to different languages or varieties" (1972: 102). Hence, the functional difference between the varieties is more crucial to Fishman (and Fasold, who agrees with Fishman) than their prestige. The only criterion which all have agreed on is function, with only slight overlapping of the social tasks assigned to each variety.

Hence, Fishman (by implication) and Fasold (explicitly) include regional and stylistic variation. But there are no real high or low varieties in such cases; speech styles do not carry the same social connotations that true "high" or "low" varieties do. Both Fishman and Fasold's views trivialize the notion of diglossia, since any stable situation in which two or more varieties are spoken within the same speech community would be diglossic.

My goal is not to argue for diglossia involving Sanskrit and Prākrit in terms of all characteristics stated by Ferguson; I leave that to present and future Sanskrit scholars.

3. Burrow (1973: 61) points out that such modifications (which he terms [false] Sanskritization) abound in Buddhist Hybrid Sanskrit, which is essentially a Sanskritized Prākrit. Here, many Prākrit words are modified to take on Sanskrit patterns, as opposed to merely substituting the equivalent Sanskrit word. E.g. Pkt. bhikkhussa, gen. sg. of bhikkhu 'monk' is changed to bhiksusya, thus "undoing" the changes from Sanskrit to Prākrit. (Skt. sy become s in Pkt. as in Skt. tasya, Pkt. tassa 'his'; cf. also the Sanskrit equivalent form bhiksos.) The discussion centers only on changes in Vedic and Classical Sanskrit, although the evidence from Buddhist Sanskrit does not detract from the argument.

4. Mayrhofer is less certain of the origin of some forms than of others (*vielleicht Hypersanskritismus*). With the exception of some forms which Mayrhofer explicitly stated could not be hyper-Sanskritisms, I considered any form that could be a hypercorrection to be an actual hyper-Sanskritism.

5. Andronov (1977) invokes hypercorrection as an explanation for certain morphological changes in Dravidian. (Only one of his examples is an actual hypercorrection; the remaining appear to be due to folk etymology or analogy.) The colloquial varieties of Tamil and Malayalam show an alternation in roots between i/e and between u/o, with the high vowels lowered to their mid counterparts when the vowel in the following syllable is a. Literary Tamil and Malayam, however, show no alternation; only i and u occur under this condition. Earlier scholars have disregarded these facts because they would involve the following sequence of events: first, Proto-South-Dravidian contained high vowels which were lowered before a syllable containing a. Then these mid vowels were raised in Proto-Tamil-Malayalam, followed by lowering in colloquial Tamil and Malayam, but not in the literary forms. However, there is no motivation for such a chronology.

According to Andronov, hypercorrection is the only logical explanation. Vowel lowering occurred occurred only once, in Proto South Dravidian. Educated Tamil and Malayam speakers felt that such lowering was "incorrect" Tamil. In the early stages of Tamil, e and o could occur before syllables containing a which were not derived from i or u, but which were originally mid vowels. Speakers of what came to be known as literary Tamil (the high variety) retained the original high vowels before a, and raised the original mid vowels before a so as to not sound like speakers of the colloquial variety.

6. This list is not by any means a complete list of hyper-Sanskritisms.

7. I have been unable to find any attested Sanskrit form containing *vv*. This is somewhat unusual because *y*, which, like *v*, is a semi-vowel, can occur as a geminate (*śayyāsanabhogāḥ*, 'lying, sitting, and eating'). But there are situations in which *vv* could potentially occur. Whitney (section 228) mentions that consonants (except for spirants preceding vowels) could optionally (and sometimes obligatorily) be geminated after *r* (and, for some grammarians, *h*, *l*, or *v*). Citing Huck and Pandharipande (p. 116), Brian Joseph pointed out to me that gemination in taunts was prescriptively incorrect (*putrādinī*, not *puttrādinī* 'cruel mother'), implying that Sanskrit speakers *did* geminate consonants in such forms. Also, two secondary sources (Coulson 1976: 24 and Kale 1969: 10) give ligatures for *vv*. However, they cite no forms containing this sequence; perhaps these ligatures are hypothetical.

References

Ananthanarayana, H. S. (1973). A Prākrit Reader. Central Institute of Indian Languages: Mysore.

Andronov, M. (1977). "Hypercorrection in Dravidian". Indian Linguistics 38: 221-226.

Balbir, J. K. (1962). "Sanskrit Drama and the Spectators". Indo-Iranian Journal 6: 38-44.

Bloomfield, Maurice and Edgerton, Franklin. (1932). Vedic Variants, Volume II. Linguistic Society of America, University of Pennsylvania: Philadelphia.

Burrow, Thomas. (1973). The Sanskrit Language (third edition). Faber and Faber: London.

Coulson, Michael. (1976). Sanskrit: An Introduction to the Classical Language. Published in U. S. by David McKay Co. (Teach Yourself Books).

DeCamp, David. (1972). "Hypercorrection and Rule Generalization". Language in Society 1: 87-90.

Deshpande, Madhav M. (1979). Sociolinguistic Attitudes in India: An Historical Reconstruction. Linguistica Extranea, Studia 5, xvi. Karoma Publishers: Ann Arbor.

De Silva, M. W. Sugathapala. (1974). "Convergence in Diglossia: The Sinhalese Situation." Contact and Convergence in South Asian Languages (edited by Franklin C. Southworth and Mahadev L. Apte). International Journal of Dravidian Linguistics.

Emeneau, Murray B. (1966). "The Dialects of Old Indo-Aryan". Ancient Indo-European Dialects (ed. by H. Birnbaum and J. Puhvel) 123-138. University of California Press: Berkeley.

Fasold, Ralph. (1984). The Sociolinguistics of Society, Ch. 2. Basil Blackwell.

Ferguson, Charles A. (1959). "Diglossia". Word 15: 325-340.

Fishman, Joshua A. (1972). The Sociology of Language, Section VI. Newbury House: Rowley, Mass.

Gair, James W. (1968). "Sinhalese Diglossia". Anthropological Linguistics 10.8 1-15.

Hock, Hans Henrich and Pandharipande, Rajeshwari. (1976). "The Sociolinguistic Position of Sanskrit in Pre-Muslim South Asia". Studies in Language Learning, Vol. I, Number II (Spring).

Kale, Moreshvar Ramchandra. (1969). A Higher Sanskrit Grammar For the Use of Schools and Colleges. Motilal Banarsidass.

Keith, Arthur Berriedale. (1924). The Sanskrit Drama. Oxford Press.

Labov, William. (1972). Sociolinguistic Patterns, Ch. 5. University of Pennsylvania: Philadelphia.

Mayrhofer, Manfred. (1956). Kurzgefasstes Etymologisches Wörterbuch des Altindischen, Vol. I. Carl Winter: Heidelberg.

Monier-Williams, Sir Monier. (1899). A Sanskrit-English Dictionary. Oxford Press.

Pischel, Richard. (1965). Comparative Grammar of the Prākrit Languages. Translated from the German by Subhadra Jhā, second edition. Motilal Banarsidass.

Sullivan, James P. (1980). "The Validity of Literary Dialect: Evidence from the Theatrical Portrayal of Hiberno-English Forms". Language in Society 9: 195-219.

Tagare, Ganesh Vasudev. (1948). Historical Grammar of Apabhraṁśa. Deccan College Dissertation Series 5. Poona.

Whitney, William Dwight. (1889). Sanskrit Grammar, fifteenth issue. Harvard University Press: Cambridge, Mass.